

\*\*\*\*\*  
 Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd  
 \*\*\*\*\*  
 (TM)

\*\*\*\*\*  
 Release 3.1a John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd  
 \*\*\*\*\*

MPearch\_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Dec 27 11:37:08 1999; MasPar time 2173.47 Seconds

Tabular output not generated. 1374.544 Million cell updates/sec

Title: >US-09-415-540-2

Description: (1-1275) from US09415540.seq

Perfect Score: 1274 1 CAGAGGTTGGGGCTCTCT.....GAATATAATTTTTCGTG 1275

N.A. Sequence: 1 GTTCTCCAAACCCCGAGAGA.....CTTATTATAATAAACGAC

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2883791 segs, 1171580779 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: emb1-est18

1:em\_est10 2:em\_est11 3:em\_est17 4:em\_est18 5:em\_est2

6:em\_est9 7:em\_gss1

genbank-est111

8:gb\_est1 9:gb\_est10 10:gb\_est11 11:gb\_est12 12:gb\_est13

13:gb\_est14 14:gb\_est15 15:gb\_est16 16:gb\_est17

17:gb\_est18 18:gb\_est19 19:gb\_est20 20:gb\_est21

21:gb\_est22 22:gb\_est23 23:gb\_est24 24:gb\_est25

25:gb\_est26 26:gb\_est27 27:gb\_est28 28:gb\_est29

29:gb\_est30 30:gb\_est31 31:gb\_est32 32:gb\_est33 33:gb\_est34

34:gb\_est35 35:gb\_est36 36:gb\_est37 37:gb\_est38 38:gb\_est39

39:gb\_gss3 40:gb\_gss4 41:gb\_gss5 42:gb\_gss6

Statistics: Mean 11.461; Variance 2.065; scale 5.550

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	710	55.7	751	22	A1078153	0.00e+00
2	626	49.1	639	17	AA694204	0.00e+00
3	596	46.8	704	28	A1565600	0.00e+00
4	571	44.8	646	18	AA766473	0.00e+00
5	567	44.5	570	24	A1215898	0.00e+00
6	541	42.5	564	14	AA493425	0.00e+00
7	529	41.5	540	20	AA932061	0.00e+00
8	523	41.1	530	18	AA777776	0.00e+00
9	520	40.8	530	18	AA777776	0.00e+00

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	A1078153	751 bp mRNA	EST	01-OCT-1998									
	0230C03.x1	Soares_feta	0.00e+00										
	IMAGE:1676836	3' similar to SW:IPR_BOVIN P37980 INORGANIC											
	PIRPHOSPHATASE	1	GI:3412561										
	A1078153	93412561											
	A1078153	GI:3412561											
	human												
	human sapiens												
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;												
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.												
	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap											
	1. (bases 1 to 751)												
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),												
	Tumor Gene Index												
	Unpublished (1997)												
	On Jan 19, 1998 this sequence version replaced gi:2282347.												

#### ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	A1078153	751 bp mRNA	EST	01-OCT-1998									
	0230C03.x1	Soares_feta	0.00e+00										
	IMAGE:1676836	3' similar to SW:IPR_BOVIN P37980 INORGANIC											
	PIRPHOSPHATASE	1	GI:3412561										
	A1078153	93412561											
	A1078153	GI:3412561											
	human												
	human sapiens												
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;												
	Eutheria; Primates; Catarrhini; Hominiidae; Homo.												
	NCI-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap											
	1. (bases 1 to 751)												
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),												
	Tumor Gene Index												
	Unpublished (1997)												
	On Jan 19, 1998 this sequence version replaced gi:2282347.												

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Insert length: 2041 Std Error: 0.00  
 Seq primer: 40m13 fwd. ET from Amersham  
 High quality sequence stop: 452.  
 Location/Qualifiers  
 1. 751  
 /organism="Homo sapiens"  
 /note="Vector: p773D-Pac (Pharmacia) with a modified  
 polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from pooled 8-9 week

## RESULT

Db      122 AAAAGTATCTTAGTTGAGATAAAAAATGCTTAGATGCATAACATTTCGAGTATATT 181  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Cp    1155 AAAAGTATCTTAGTTGAGATATGAANAATCGCTTAGATGGATAACATTTCGAGTATATT 1096

Db 182 GGATGATCAGACAGCAATTTACTTACTAGTAGAGTCTTACAAATTTAAAGCTTTGAA 241  
 Cp 1095 GGATTGTCAGACAGCAATTTACTTACTAGTAGAGTCTTACAAATTTAAAGCTTTGAA 1036  
 Db 242 AAGCTACTACTTTTACTTCTTAATACATCCAGATGAACAGAGTAGCAGATATCAGCTTGT 301  
 Cp 1035 AAGCTACTACTTTTACTTCTTAATACATCCAGATGAACAGAGTAGCAGATATCAGCTTGT 976  
 Db 302 ATTCCAGAGAAATCTATAGTCTTCTGCTGATGAGAACCACTTATCCAGCTGTGGT 361  
 Cp 975 ATTCCAGAGAAATCTATAGTCTTCTGCTGATGAGAACCACTTATCCAGCTGTGGT 916  
 Db 362 ACTGTCAGAGCAATTCACAGAGGTGGTGAAGATCCAAATGGCTGCGAGCATCA 421  
 Cp 915 ACTGTCAGAGCAATTCACAGAGGTGGTGAAGATCCAAATGGCTGCGAGCATCA 856  
 Db 422 GGATCAGACTTGAAGGGGCTTCAGACAAAGTTGATTCATGCACTGATCTCTTTTCCA 481  
 Cp 855 GGATCAGACTTGAAGGGGCTTCAGACAAAGTTGATTCATGCACTGATCTCTTTTCCA 796  
 Db 482 TTGCTTTTCTAGTACTATAGCTTTCATGATGATGATGATGATGATGATGATGATG 541  
 Cp 795 TTGCTTTTCTAGTACTATAGCTTTCATGATGATGATGATGATGATGATGATGATG 736  
 Db 542 GGAAGTCTCTTACTTAAATCTGATTAAGGCAAACTCTTCTGCTTTTACATCA 601  
 Cp 735 GGAAGTCTCTTACTTAAATCTGATTAAGGCAAACTCTTCTGCTTTTACATCA 676  
 Db 602 GGAACCTTATACCTTCTAAACCCAGTACAGTACAGTCTTC 639  
 Cp 675 GGAACCTTATACCTTCTAAACCCAGTACAGTACAGTCTTC 638

RESULT 3  
 LOCUS A1565600 704 bp mRNA EST 26-MAR-1999  
 DEFINITION t016f09.x1 NCI\_CGAP Ut2 Homo sapiens CDNA clone IMAGE:2179241.3  
 similar to SW:IPR\_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA  
 sequence.  
 ACCESSION A1565600  
 NID 94524057  
 VERSION A1565600.1 GI:4524057  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 704)  
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3137533.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emerit-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNT at:  
 www.bio.llnl.gov/dbp/image/image.html

FEATURES  
 source  
 Seq primer: -40UP from G1bco  
 High quality sequence stop: 404.  
 Location/Qualifiers

1..704  
 /organism="Homo sapiens"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI;  
 Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT;  
 Average insert size 1.85 kb. Life Technologies catalog #:

11539-012"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2179241"  
 /clone\_lib="NCI\_CGAP\_Ut2"  
 /tissue\_type="moderately-differentiated endometrial  
 adenocarcinoma, 3 pooled tumors"  
 /lab\_host="DH10B"  
 BASE COUNT 205 a 137 c 120 g 238 t 4 others  
 ORIGIN

Query Match 46.8%; Score 596; DB 28; Length 704;  
 Best Local Similarity 97.0%; Pred. No. 0.00e+00;  
 Matches 653; Conservative 0; Mismatches 12; Indels 8; Gaps 8;

Db 13 CAGCAAAATTAATTTTATTTCTTACATATGATATATATATATATATATATATATATATAT 72  
 Cp 1275 CAGCAAAATTAATTTTATTTCTTACATATGATATATATATATATATATATATATATAT 1216  
 Db 73 TGCACATCTATTCACAGAGGACTTCCAAATGGCACTGTTGATATTTAAGCATGTGCT 132  
 Cp 1215 TGCACATCTATTCACAGAGGACTTCCAAATGGCACTGTTGATATTTAAGCATGTGCT 1156  
 Db 133 AAGAGTTATCTTATGATATATGAAATATGCTTATGATGATATATATATATATATATATAT 192  
 Cp 1155 AAGAGTTATCTTATGATATATGAAATATGCTTATGATGATATATATATATATATATATAT 1096  
 Db 193 GGATTTGTCACAGACAAATTTACTTATGATGATGATGATGATGATGATGATGATGATG 252  
 Cp 1095 GGATTTGTCACAGACAAATTTACTTATGATGATGATGATGATGATGATGATGATGATG 1036  
 Db 253 AAGCTACTCTTACTTCTTAATCATCCAGATGAACAGATGATGATGATGATGATGATGATG 312  
 Cp 1035 AAGCTACTCTTACTTCTTAATCATCCAGATGAACAGATGATGATGATGATGATGATGATG 976  
 Db 313 ATCCAGAGAAATCTATAGTCTTCTGCTGATGAGAACCACTTATCCAGCTGTGGT 372  
 Cp 975 ATCCAGAGAAATCTATAGTCTTCTGCTGATGAGAACCACTTATCCAGCTGTGGT 916  
 Db 373 ACTGTCAGAGCAATTCACAGGGGTGTGTTAAGCATCCAAATGGCTCTGCGAGCATCA 432  
 Cp 915 ACTGTCAGAGCAATTCACAGGGGTGTGTTAAGCATCCAAATGGCTCTGCGAGCATCA 856  
 Db 433 GGATCAGACTTGAAGGGGCTCTAGACAAAGTGTATATGCACTGATCTCTTTTCCA 492  
 Cp 855 GGATCAGACTTGAAGGGGCTCTAGACAAAGTGTATATGCACTGATCTCTTTTCCA 796  
 Db 493 TTGCTTTTCTAGTACTATAGCTTTCATGATGATGATGATGATGATGATGATGATGATG 551  
 Cp 795 TTGCTTTTCTAGTACTATAGCTTTCATGATGATGATGATGATGATGATGATGATGATG 736  
 Db 552 GCAA-GTCTTATCTTTAA-TTCTGCAT-AAAGCAAACTATTTTCTGTTT-CCATCA 607  
 Cp 735 GCAAGTCTTATCTTTAAATTTCTGCATTAAGCAAACTATTTTCTGTTTCCATCA 676  
 Db 608 AGAAGCTTATACCTTCTTAAC-AGNNACAGTAGCCTCTAGTAGCCAGAGTTTCA-CCG 665  
 Cp 675 GGAAGCTTATACCTTCTTAAC-AGNNACAGTAGCCTCTAGTAGAG-CCAGGTTTCAAGCCG 617  
 Db 666 TTGACATCATTTG 678  
 Cp 616 TTGACATCATTTG 604

RESULT 4  
 LOCUS AA766473 646 bp mRNA EST 28-JAN-1998  
 DEFINITION ca339g06.s1 NCI\_CGAP\_GCB1 Homo sapiens CDNA clone IMAGE:1307002.3  
 similar to SW:IPR\_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA  
 sequence.  
 ACCESSION AA766473  
 NID 92817711  
 VERSION AA766473.1 GI:2817711  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 646)

REFERENCE  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL  
On Jan 19, 1998 this sequence version replaced gi:2045917.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbtp/image/image.html](http://www.bio.llnl.gov/dbtp/image/image.html)

Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 190.

FEATURES  
Location/Qualifiers

1..646

/organism="Homo sapiens"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
geminal center B cells by flow sorting (CD20<sup>+</sup>, IgD<sup>-</sup>),  
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman  
(NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was  
primed with a Not I - oligo(dT) primer.  
[5'-TGTTCCAACTGCAAGTGGAGCGGCGCTTTTATTTTATTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

/db\_xref="taxon:9606"  
/clone\_id="NCI CGAP GCBI"  
/tissue\_type="geminal center B cell"  
/lab\_host="DH08"

BASE COUNT 180 a 121 c 108 g 237 t

ORIGIN

\*Query Match 44.8%; Score 571; DB 18; Length 646;

Best Local Similarity 97.2%; Pred. No. 0.00e+00;  
Matches 588; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Db 42 CAGCAAAATATTTATTTCTTACATATGTATACATATGATGCTCCCT 101  
Cp 1275 CAGCAAAATATTTATTTCTTACATATGTATACATATGATGCTCCCT 1216  
Db 102 TGCACATCTATTCACAGTCTTCCAAATGACAACTGCTTGAATTTAAGATGCT 161  
Cp 1215 TGCACATCTATTCACAGTCTTCCAAATGACAACTGCTTGAATTTAAGATGCT 1156  
Db 162 AAAGTATCTAGTATGATATGAAAAATGCTTAGATGATTAACATTCGATATAT 221  
Cp 1155 AAAGTATCTAGTATGATATGAAAAATGCTTAGATGATTAACATTCGATATAT 1996  
Db 222 GGATTAAGTATGATATGATATGATGATGATGATGATGATGATGATGATGAT 281  
Cp 1095 GGATTAAGTATGATATGATATGATGATGATGATGATGATGATGATGATGAT 1036  
Db 282 AAGCTACTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 341  
Cp 1035 AAGCTACTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACTTACT 976

Db 342 ATTCCAGGAATCTCATTTAGTTTCTGATGATGATGATGATGATGATGATGATGAT 401  
Cp 975 ATTCCAGGAATCTCATTTAGTTTCTGATGATGATGATGATGATGATGATGATGAT 916  
Db 402 ACTGTGAGGAGATTCACAGGAGTGTGTAAAGCATTCACATGCTTGGACATCA 461  
Cp 915 ACTGTGAGGAGATTCACAGGAGTGTGTAAAGCATTCACATGCTTGGACATCA 856  
Db 462 GGATTCACACTTGAGGGGCTCTCAGACAGGTGTATTCATGCACTGATTCCTTCC 521  
Cp 855 GGATTCACACTTGAGGGGCTCTCAGACAGGTGTATTCATGCACTGATTCCTTCC 796  
Db 522 TTGCTTTCTAGGCGCATTAAGCTTCCAAATGATGATGATGATGATGATGATGAT 581  
Cp 795 TTGCTTTCTAGGCGCATTAAGCTTCCAAATGATGATGATGATGATGATGATGAT 736  
Db 582 GCAAGTCTTATCTTAAATTCGATTAACCCCAACATCATTTGCTTCCATCA 641  
Cp 735 GCAAGTCTTATCTTAAATTCGATTAACCCCAACATCATTTGCTTCCATCA 676  
Db 642 GGAGC 646  
Cp 675 GGAGC 671

## RESULT

5 A121898 570 bp mRNA EST 21-DEC-1998

LOCUS  
DEFINITION  
gms3509.x1 NCI-CGAP L05 Homo sapiens cDNA clone IMAGE:1883873 3'  
similar to SW:IPYR\_BOVIN P37980 INORGANIC PYROPHOSPHATASE ;, mRNA  
sequence.

ACCESSION  
NID  
VERSION  
KEYWORDS  
SOURCE

A121898  
93784939  
A121898.1 GI:3784939  
EST.  
human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
1 (bases 1 to 570)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL  
On Jan 14, 1998 this sequence version replaced gi:1796596.

COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbtp/image/image.html](http://www.bio.llnl.gov/dbtp/image/image.html)

Insert Length: 1001 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 353.

FEATURES  
Location/Qualifiers

1..570

/organism="Homo sapiens"  
/note="Organ: Lung; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; 1st strand cDNA was prepared from  
neuroendocrine lung carcinoid, and was then primed with a  
Not I - oligo(dT) primer. Double-stranded cDNA was ligated  
to Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library is normalized. Library was  
constructed by Bento Soares and M. Fatima Bonaldo."  
/db\_xref="taxon:9606"  
/clone="IMAGE:1883873"







Query Match 41.5%; Score 529; DB 36; Length 638;  
Best Local Similarity 96.6%; Pred. No. 0.00e+00;  
Matches 576; Conservative 0; Mismatches 14; Indels 6; Gaps 6

	DB	Accession	Score	Length	Mismatches	Indels	Gaps
Db	2	CAGCAAAATTAATTTTATTTCCTACATATGGTAAACATATACATCCAAATATGTGCTCCCT	61				
Db	1275	CAGCAAAATTAATTTTATTTCCTACATATGGTAAACATATACATCCAAATATGTGCTCCCT	1216				
Db	62	TGCACATCATCACAAGTGACTGCTCCAAATATACAACTGCTTTGATTTTAAAGCATGTGCT	121				
Db	1215	TGCACATCATCACAAGTGACTGCTCCAAATATACAACTGCTTTGATTTTAAAGCATGTGCT	1156				
Db	122	AAAAGTTATCTTAGTTGAGATATGAATAAATGCTTTAGATGATTAACATCTCGACTATAT	181				
Db	1155	AAAAGTTATCTTAGTTGAGATATGAATAAATGCTTTAGATGATTAACATCTCGACTATAT	1096				
Db	182	GGATTAGTCACAGACAGAATTTACTTTAGTAGATGAGTTCTTACAAATTTAAACGCTTGA	241				
Db	1095	GGATTAGTCACAGACAGAATTTACTTTAGTAGATGAGTTCTTACAAATTTAAACGCTTGA	1037				
Db	242	AACCTCTACTTTTACTTCTTAATATCATCCAGATGACAGCATGTAGCATATACGCTTG	301				
Db	1036	AACCTCTACTTTTACTTCTTAATATCATCCAGATGACAGCATGTAGCATATACGCTTG	977				
Db	302	TATTCACAGAGAAATCTCATTTAGTTTTCTGSGTAGAGAACACTTATCCAGCTGTGTTG	361				
Db	976	TATTCACAGAGAAATCTCATTTAGTTTTCTGSGTAGAGAACACTTATCCAGCTGTGTTG	917				
Db	362	TACTGTGCAGGCAGATTCCACAGGGTGTTGTTAAAGCATATCCAAATGGGCTCTGCAGCAT	421				
Db	916	TACTGTGCAGGCAGATTCCACAGGGTGTTGTTAAAGCATATCCAAATGGGCTCTGCAGCAT	858				
Db	422	CAGGATCACACTTGAAGGGGCTCTCAGCAAAGTTGGTATTATGACAACTGATGCCCTT	481				
Db	857	CAGGATCACACTTGAAGGGGCTCTCAGCAAAGTTGGTATTATGACAACTGATGCCCTT	799				
Db	482	CCAATCGMTTCCTAGTACATTAATGCTTCCAAATGGGCTAGTGGCCTTAATAATATCA	541				
Db	798	CCAATCGMTTCCTAGTACATTAATGCTTCCAAATGGGCTAGTGGCCTTAATAATATCA	739				
Db	542	ATGGGCAAAAGTCCTATCCCTTAATATCNGCACTTAA-CGCA-CTCATTTCCGAGCTT	595				
Db	738	ATGG-CAAAGCTCTATCTTAAATTCGTGATTTAAAGCAAACTCATTTCTGCTT	684				
RESULT	9	A1219448	572 bp	mRNA	EST	30-NOV-1998	
JOCUS		gnl3c03.x1 Soares_NFL.T.GBC.S1 Homo sapiens CDNA clone					
DEFINITION		IMAGE:1844548.3 similar to SW:PYPY_BOVIN P37980 INORGANIC					
		PYROPHOSPHATASE ; mRNA sequence.					
ACCESSION		A1219448					
NID		g3801651					
VERSION		A1219448.1 GI:3801651					
KEYWORDS		EST.					
SOURCE		human.					
ORGANISM		Homo sapiens					
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;					
AUTHORS		Eutheria; Primates; Catarrhini; Homidae; Homo.					
TITLE		(bases 1 to 572)					
		NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap.					
		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),					
JOURNAL		Tumor Gene Index					
		Unpublished (1997)					

On Aug 21, 1998 this sequence version replaced

```

/organism="Homo sapiens"
/notice="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-GCAP-GCB1) were mixed, and ss circles were made in
vivo. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The drive
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries (The pools consisted of
I.M.A.G.E. clones 297480-302087, 682652-687259,
726408-728711, and 729096-731359. Subtraction by Bento
Soares and M. Fatima Bonaldo. "
/db_xref="taxon:9606"
/clone="IMAGE:1844548"
/clone_lib="Soares-NFL-T-GBC-S1"
/lab_host="DH10B"
1173 a 109 c 97 g 192 t 1 others

```

Query Match	41.1%	Score 523	DB 24	Length 572
Best Local Similarity	97.6%	Pred. No. 0.00e+00		
Matches 561	Conservative	0	Mismatches 9	Indels 5
			Gaps 5	
Db	3	CAGCAAAATAATTTATTTCTCTTAACATATGTGTAACATATACATCATATATGTCTCCCT	62	
Cp	1275	CAGCAAAATAATTTATTTCTCTTAACATATGTGTAACATATACATCATATATGTCTCCCT	1216	
Db	63	TGCACATCTATTCACAAAGTGACTTCCAAATGACACACTGCTTTGATATTTAAGCATGTCT	122	
Cp	1215	TGCACATCTATTCACAAAGTGACTTCCAAAGACAACTGCTTTATATTTAAGCATGTGCT	1156	
Db	123	AAAAGTTATCTTAAGTATGATGATGAAATATGCTTTGATGATGATGATGATGATGAT	182	
Cp	1155	AAAAGTTATCTTAAGTATGATGATGAAATATGCTTTGATGATGATGATGATGATGAT	1096	
Db	183	GGATTAGTCACAGCAGAAATTACTTTAGTAGATGAGTTCTACAAATTTAAAGCTTTGAA	242	
Cp	1095	GGATTAGTCACAGCAGAAATTACTTTAGTAGATGAGTTCTACAAATTTAAAGCTTTGAA	1036	
Db	243	AAGCTACTACTTTACTTCTTAATATCATCCAGATGACACATGATAGCAATATACGTTGT	302	
Cp	1035	AAGCTACTACTTTACTTCTTAATATCATCCAGATGACACATGATAGCAATATACGTTGT	976	
Db	303	ATTCAGAGAAATCTCATTAAGTTTCTGGATGAGAAACACTTATCCAGCTGTGTTGT	362	
Cp	975	ATTCAGAGAAATCTCATTAAGTTTCTGGATGAGAAACACTTATCCAGCTGTGTTGT	916	
Db	363	ACTGTGACAGCAGATTCACAGGGGTGTG-G-AAAGCATCCCAATGGCTGTGGCAGCATCA	421	
Dp	915	ACTGTGACAGCAGATTCACAGGGGTGTGTAAGCATCCCAATGGCTGTGGCAGCATCA	856	
Db	422	GGATTCACACTTGAAGGGGCTCTCAGACAAAGTTGTATATGCACTGATCTCTTTCCA	481	
Cp	855	GGATTCACACTTGAAGGGGCTCTCAGACAAAGTTGTATATGCACTGATCTCTTTCCA	796	
Db	482	T-GGTTTTCTTA-TCATTAATGGCTTCCAGTTCNCATGA-TGGTTTTA-TAATATAAG	537	
Cp	795	TTGGTTTTCTTAAGTACATTAATGGCTTCCAGTTCNCATGAAGTGGCTTTTAATATATCAATG	736	
Db	538	GCAAGTCCTTTATCTTAAATTCATGCAATAAAGCG	572	

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewartson.wustl.edu  
 This clone is available royalty-free through LINL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 271.  
 Location/Qualifiers  
 1. 530

```
Query Match      40.8%   Score 520; DB 18; Length 530;
Best Local Similarity 99.1%   Pred. No. 0.00e+00;
Matches 525, Conservative 0; Mismatches 5; Indels 0; Gaps 0
```

Db	121	AAAGTATCTTACTAGATATGAAAAATGCTTTAGATGGATTAACATCTGAGTATAT	180
Cp	1155	AAAAGTTATCTTACTAGATATGAAAAATGCTTTAGATGGATTAACATCTGAGTATAT	10986
Db	181	GGATTAGCACAGAGAAATTTACTTATAGTAGTGCTCTCAAAATTTAAAGCTTTGAA	240
Cp	1095	GGATTAGTCACAGAGAAATTTACTTATAGTAGTGCTCTCAAAATTTAAAGCTTTGAA	10366
Db	241	AAGTACTACTTTTACTCTTAATACATCCAGATGAAACAGATGTGACATATCACCTTGT	300
Cp	1035	AAGTACTACTTTTACTCTTAAATACATCCAGATGAAACAGATGTGACATATCACCTTGT	976
Db	301	ATTCACAGAAATCTAATTTGTTTCTGTGATGATGACCACTTTTCCAGCTGTGTTGT	360
Cp	975	ATTCACAGAAATCTAATTTGTTTCTGTGATGATGACCACTTTTCCAGCTGTGTTGT	916
Db	361	ACTGTGACAGGCAGATTCACAGGGTGTGTAAAGCATCCACAAATGGCTGTGGACATCA	420
Cp	915	ACTGTGACAGGCAGATTCACAGGGTGTGTAAAGCATCCACAAATGGCTGTGGACATCA	856
Db	421	GGATCAACTTGAAGGGGCTCTACAGCAAAGTTGTAATTCATGCAACTGATTCCTTTTCA	480
Cp	855	GGATCAACTTGAAGGGGCTCTACAGCAAAGTTGTAATTCATGCAACTGATTCCTTTTCA	796
Db	481	ATCGATTCTCTACTACTATATGCTTCCAAATGATATAGTATAGGCTTTAT	530
Cp	795	ATCGATTCTCTACTACTATATGCTTCCAAATGATATAGTATAGGCTTTTAT	746

LOCUS 11  
DEFINITION 20-ANG-1998  
gb91e02.x1 Soares-fetal\_heart\_NbH19M Homo sapiens CDNA clone  
IMAGE:1707482.3' similar to SW:IPYR\_BOVIN P37980 INORGANIC  
PYROPHOSPHATASE ; , mRNA sequence.  
ACCESSION A1096370 521 bp mRNA  
NID 93446281  
VERSION A1096370.1 GI:3446281  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 521)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
COMMENT Tumor Gene Index  
Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1400967.

CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL : contact the  
IMAGE Consortium ([infoimage@lnl.gov](mailto:infoimage@lnl.gov)) for further information.  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence, stop: 452.

FEATURES  
source 1..521  
Location/Qualifiers

```

/organism="Homo sapiens"
/notice="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer
TGGTACCAATCTGACGTGGAGCGGCCGCTTTTCTTTTCTTTTCTT 3'),
double-stranded cDNA was size selected, ligated to Eco
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Patina Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH119."
/db_xref="taxon:9606"
/clone="IMAGE:1707482"

```



/clone\_lib="Soares\_fetal\_heart\_NbH19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 BASE COUNT 156 a 99 c 93 g 173 t  
 ORIGIN

Query Match 40.7%; Score 519; DB 22; Length 521;  
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;  
 Matches 520; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 1 CAGCAAAATATTTATTTCTTACATATGTATACATATACATATCAATATGTCCTCCCT 60  
 Cp 1275 CAGCAAAATATTTATTTCTTACATATGTATACATATACATATCAATATGTCCTCCCT 1216  
 Db 61 TGCACATCTATTCACAGAGCTCCCAATGACAACTGCTTGTATTTAAGCATGTGCT 120  
 Cp 1215 TGCACATCTATTCACAGAGCTCCCAATGACAACTGCTTGTATTTAAGCATGTGCT 1156  
 Db 121 AAAATCTTCTAGTGTGATGATGAAATGCTTTAGATGATGATGATGATGATGATGAT 180  
 Cp 1155 AAAATCTTCTAGTGTGATGATGAAATGCTTTAGATGATGATGATGATGATGATGAT 1096  
 Db 181 GGATAGTACAGCAGCAATTTACTTTAGTATGATGATGATGATGATGATGATGATGAT 240  
 Cp 1095 GGATAGTACAGCAGCAATTTACTTTAGTATGATGATGATGATGATGATGATGATGAT 1036  
 Db 241 AAGTACTACTTTTACTTCTTATACATCCAGATGACAGATGATGATGATGATGATGATGAT 300  
 Cp 1035 AAGTACTACTTTTACTTCTTATACATCCAGATGACAGATGATGATGATGATGATGATGAT 976  
 Db 301 ATTCAGAGAAATCTCATTTAGTGTGATGATGATGATGATGATGATGATGATGATGAT 360  
 Cp 975 ATTCAGAGAAATCTCATTTAGTGTGATGATGATGATGATGATGATGATGATGATGAT 916  
 Db 361 ACTGTGACAGCAGATTTCACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 420  
 Cp 915 ACTGTGACAGCAGATTTCACTGCTGCTGATGATGATGATGATGATGATGATGATGAT 856  
 Db 421 GGATACACTTGAAGGGCTCTCAACAAAGTTGATTCATCACTGATGATGATGATGATGAT 480  
 Cp 855 GGATACACTTGAAGGGCTCTCAACAAAGTTGATTCATCACTGATGATGATGATGATGAT 796  
 Db 481 TTCGTTTCTTGTACATTAATGCTTTCCATGATGATGATGATGATGATGATGATGAT 521  
 Cp 795 TTCGTTTCTTGTACATTAATGCTTTCCATGATGATGATGATGATGATGATGATGAT 755

RESULT 12 AA314881 528 bp mRNA EST 19-APR-1997  
 LOCUS EST186922 HCC cell line (metastasis to liver in mouse) II Homo  
 DEFINITION sapiens cDNA 5' end similar to similar to inorganic  
 pyrophosphatase, mRNA sequence.  
 ACCESSION AA314881  
 MID 91967451  
 VERSION AA314881.1 GI:1967451  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 528)  
 AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fulmer,R.A.,  
 Bull,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,  
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-  
 Fitting,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,  
 Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,  
 Glodet,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,  
 Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaro,S.M., Merrick,J.M.,  
 Moreno-Palanges,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,  
 Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,  
 Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y.,  
 Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT  
 On Apr 14, 1993 this sequence version replaced gi:785883.  
 Other ESTs: THC195084  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi.html>)  
 Seq primer: M13 Reverse.

FEATURES  
 Location/Qualifiers  
 1..528  
 /organism="Homo sapiens"  
 /note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
 XhoI"  
 /db\_xref="ATCC (inhost):113250"  
 /db\_xref="taxon:9606"  
 /map="844D06; 5; 5q31.1-5q31.3"  
 /clone\_lib="HCC cell line (metastasis to liver in mouse)  
 II"  
 /tissue\_type="colon"  
 /cell\_type="KM12SM"  
 /cell\_line="KM12C(HCC)metastasis into mouse (liver)"

BASE COUNT 155 a 101 c 140 g 132 t  
 ORIGIN

Query Match 40.7%; Score 518; DB 11; Length 528;  
 Best Local Similarity 99.4%; Pred. No. 0.00e+00;  
 Matches 526; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 1 TGGTGGCTGTGTGCGAGCGGGCGGCGAGACTCCGCGCATATGAGCGGCTTCACACCG 60  
 Cp 48 TGGTGGCTGTGTGCGAGCGGGCGGCGAGACTCCGCGCATATGAGCGGCTTCACACCG 107  
 Db 61 AGGAGCGGGCGGCGGCGGCTTCTCCCTGAGATACGAGTCTCTCTCAAAATGGAAGAGC 119  
 Cp 108 AGGAGCGGGCGGCGGCGGCTTCTCCCTGAGATACGAGTCTCTCTCAAAATGGAAGAGC 167  
 Db 120 AATATATCTTCCATTTTCATGATGATTCATTCATTCATTCATTCATTCATTCATTCAT 179  
 Cp 168 AATATATCTTCCATTTTCATGATGATTCATTCATTCATTCATTCATTCATTCATTCAT 227  
 Db 180 TAGTGAAGTACACGCTGTGTCTAATGCAAAATGAGATGCTCAAGAGACCTTTAA 239  
 Cp 228 TAGTGAAGTACACGCTGTGTCTAATGCAAAATGAGATGCTCAAGAGACCTTTAA 287  
 Db 240 ACCCTATTAACAAATGTAAGAAAAAGAAAACTTCGCTATGTCGAAATTTGTTCCCT 299  
 Cp 288 ACCCTATTAACAAATGTAAGAAAAAGAAAACTTCGCTATGTCGAAATTTGTTCCCT 347  
 Db 300 ATAAAGATATATCTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 359  
 Cp 348 ATAAAGATATATCTGGAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 407  
 Db 360 ATGATTAACATCTGCTGTGTGTGTGATGATGATGATGATGATGATGATGATGATGAT 419  
 Cp 408 ATGATTAACATCTGCTGTGTGTGTGATGATGATGATGATGATGATGATGATGATGAT 467  
 Db 420 GCAAGTATGTCAGAGAGTGAATTAATGCGGTGAAGATTTAGCATATTTGCTATGCA 479

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

This clone is available royalty-free through LINT; contact the IMAG Consortium ([info@imag.llnl.gov](mailto:info@imag.llnl.gov)) for further information. Possible reversed clone similarity on wrong strand  
Seq primer: -40m3 fwd. Et from Amersham  
High quality sequence stop: 273

Query Match	40.7%	Score 518;	DB 16;	Length 572;
Best Local Similarity	97.8%;	Pred. No. 0.00e+00;		
Matches 530; Conservative	0;	Mismatches 12;	Indels 0;	Gaps 0

Db 211 GGATTAGTACAGCAGAAATTACTTACTTAGTACGTAGTGTCTCAAAATTTAAAGCTTTGGA 270  
 Cp 1095 GGATTAGTACAGCAGAAATTACTTACTTAGTACGTAGTGTCTCAAAATTTAAAGCTTTGGA 1036  
 Db 271 AAGCTACTACTTTTACTTCTTAATACATCCAGATGAGACAGATGTACCAATATACGCTGT 330  
 Cp 1035 AAGCTACTACTTTTACTTCTTAATACATCCAGATGAGACAGATGTACCAATATACGCTGT 976  
 Db 331 ATTCCAGAGAAATTCATTAGTTTGTGCTGTATGAGAACCACTTATTCACAGTCTGTGGT 390  
 Cp 975 ATTCCAGAGAAATTCATTAGTTTGTGCTGTATGAGAACCACTTATTCACAGTCTGTGGT 916  
 Db 391 ACTGTGCAGACGAGATTACACAGGCTGTGTGTTAAAGCATCCACAATGGCTGTGCAGATCA 450  
 Cp 915 ACTGTGCAGACGAGATTACACAGGCTGTGTGTTAAAGCATCCACAATGGCTGTGCAGATCA 856  
 Db 451 GGATCACACTTGAAGGGGCTCTCAGACAAAGTTGATTTCATCAACTGATTCCTTTTCCA 510  
 Cp 855 GGATCACACTTGAAGGGGCTCTCAGACAAAGTTGATTTCATCAACTGATTCCTTTTCCA 796  
 Db 511 TTCCGTTTCCCTAAGTCCTAATGCTTTCCAAATGTCGGAAGTCTTTTAAATATCAAG 570  
 Cp 795 TTCCGTTTCCCTAAGTCCTAATGCTTTCCAAATGTCGGAAGTCTTTTAAATATCAAG 736  
 Db 571 GC 572  
 Cp 735 GC 734  
 RESULT 14  
 LOCUS AA702134 521 bp mRNA EST 19-DEC-1997  
 DEFINITION Z185908.H1 Soares fetal\_liver spleen\_INFLS.S1 Homo sapiens CDNA  
 clone IMAGE:447514 3' similar to SW:IPYR\_BOVIN P37980 INORGANIC  
 PYROPHOSPHATASE ; mRNA sequence.  
 ACCESSION AA702134  
 NID 92705247  
 VERSION AA702134.1 GI:2705247  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 521)  
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,  
 Krizman,D., Kucuba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,  
 Matlin,J., Moore,B., Schellenberg,K., Stepien,M., Tan,F.,  
 Teisinger,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.  
 WashU-NCI human EST project  
 TITLE Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394456.  
 CONTACT: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL ; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40m3 fwd. RT from Amersham  
 High quality sequence stop: 399.  
 Location/Qualifiers  
 1..521

```

FEATURES
Source
Location/Qualifiers
1..521
/organism="Homo sapiens"
/note="Organ: Liver and Spleen; Vector: pTV73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI".
This is a subtracted version of the Original Soares fetal
liver spleen INFEIS library. 1st strand cDNA was primed
with a Pac I - Oligo(CT) primer [5'
AAGTCGACAGATTAAATTAAAGACCTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I

```



Tue Dec 28 11:08:20 1999

US-09-415-540-2.rst

Page 12

Db 489 ACCAGTACCACACAGACGCGTGAATAGTNTTTCATCACACAGAAACCTNATGGAGATTCC 548  
Qy 911 AC-AGTACCACAGACGCGTGAATAGTGGTTCATCACACAGAAACCTAAG-AGATTTC 968  
Db 549 CTGGGATTCACCGCTGATGATGCCACATCGTG 580  
Qy 969 CTGGATATCAAGC-TGATATTGCTACATCGTG 999

Search completed: Mon Dec 27 12:53:35 1999  
Job time : 4587 secs.

```

RESULT      3      CEC47E12      37855 bp      DNA      INV      23-NOV-1998
LOCUS
DEFINITION  Caenorhabditis elegans cosmid C47E12, complete sequence.
ACCESSION   Z68882
NID         91167468
VERSION     Z68882.1 GI:1167468
KEYWORDS    HTG: ATP carrier protein; Alpha-mannosidase; Inorganic
              pyrophosphatase; Seryl-tRNA synthetase; Ubiquitin-activating
              enzyme; uilp like; Yeast yd78 like.
SOURCE      Caenorhabditis elegans
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
              Rhabditina; Rhabditidae; Rhabditidae; Peleodermata; Caenorhabditis.
REFERENCE   1 (bases 1 to 37855)
AUTHORS     Coles, L.
TITLE       Direct Submission
JOURNAL     Submitted (25-JAN-1996) Louis, MO 63110, USA. E-mail:
REFERENCE   2 (bases 1 to 37855)
AUTHORS     Jeesander, ac.uk or rwenematode.wustl.edu
TITLE       JOURNAL
REFERENCE   3 (bases 1 to 37855)
AUTHORS     Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
              Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
              Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
              Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
              Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
              Latteille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
              O'Callaghan, M., Parsons, J., Percy, C., Ritken, L., Roopra, A.,
              Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E.,
              Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaadin, M.,
              Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
              Wilkinon-Spratt, J. and Wohldman, P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
              elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE     94150718
COMMENT     Coding sequences below are predicted from computer analysis, using
              predictions from Genefinder (P. Green, U. Washington), and other
              available information.
              For a graphical representation of this sequence and its analysis
              see:
              http://webace.sanger.ac.uk/cgi-
              bin/display?db=wormace&class=Sequence&object=C47E12
              Current sequence finishing criteria for the C. elegans genome
              sequencing consortium are that all bases are either sequenced
              unambiguously on both strands, or on a single strand with both a
              dye primer and dye terminator reaction, from distinct subclones.
              Exceptions are indicated by an explicit note.
              IMPORTANT: This sequence is NOT necessarily the entire insert of
              the specified clone. It may be shorter because we only sequence
              overlapping sections once, or longer because we arrange for a small
              overlap between neighbouring submissions.
              IMPORTANT: This sequence is not the entire insert of clone C47E12.
              It may be shorter because we only sequence overlapping sections
              once, or longer because we arrange for a small overlap between
              neighbouring submissions.
              The true left end of clone C47E12 is at 1 in this sequence. The
              true left end of clone F44D12 is at 3752 in this sequence. The
              true right end of clone T13H10 is at 4978 in this sequence. The
              start of this sequence (1..104) overlaps with the end of sequence
              Z69361.
              The end of this sequence (37752..37855) overlaps with the start of
              sequence Z68298.
FEATURES
  source
    location/Qualifiers
      1..37855
      /organism="Caenorhabditis elegans"
      /db_xref="taxon:6239"
      /chromosome="iv"
      /clone="C47E12"
      959..2262
      /gene="C47E12.11"
      join(959..1428,1483..1578,1623..1728,1787..2112,
      2160..2262)
      /gene="C47E12.11"
  gene
    cds
      join(959..1428,1483..1578,1623..1728,1787..2112,
      2160..2262)
      /gene="C47E12.11"

```

Cp	510	AGACCTTCAAGCCAAATATTTCACCGTTTCACATACTGCTTCACATTTCCATTCACACA	451
Db	27375	TCATTTGGATCATTTGTCTCCTTTTAGTCCAGTGTCAGGAACACATGTTTGATCTTCC	27434
Cp	450	TCAATTTGGGTCATTTGTGCACCACCAACACCGATGTTTATCATTTGTGCCCTGGGCTTCCC	391
Db	27435	CATCTCTGTGGAGAAGTCGCTTAATTCAGATGATGATCTTGTCTGGGAAAATTTGTGA	27494
Cp	390	CAACTCTAGAGGATGGACACCATATCTTCAGATATATCTTTAATCGGGAACAAATTTGCA	331
Db	27495	ACAAAGCGAGCTACACCCCTTCTCTCATCTGCTTGATGGCGAATGCTCTTGTTG	27554
Cp	330	ACATAGGGAAGATTTCCTTTTTCACATCTTGTTTAATAGGGTTAAAGGCTCCTTTGTA	271
Db	27555	GCAATTTCCATTTTAGCATTTGTCACACGGGATATCCATCATCATGT	27603
Cp	270	GCATCTCCATTTTTCATTTAGACACGCGTGTTACTTCACTACCATCATGT	222
RESULT	4		
LOCUS	AF085601	873 bp	mRNA
DEFINITION	Drosophila melanogaster	inorganic pyrophosphatase Nurf-38	(Nurf-38)
ACCESSION	AF085601		complete cds.
NID	93600093		
VERSION	AF085601.1	GI:3600093	
KEYWORDS	fruit fly,		
SOURCE	Drosophila melanogaster		
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Diptera; Brachyera; Muscomorpha; Ephydroidea;		
	Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 873)		
AUTHORS	Gdula,D.A., Sandlitopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.		
TITLE	Inorganic pyrophosphatase is a component of the Drosophila		
JOURNAL	nucleosome remodeling factor complex		
MEDLINE	Genes Dev. 12 (20), 3206-3216 (1998)		
REFERENCE	2 (bases 1 to 873)		
AUTHORS	Gdula,D.A., Sandlitopoulos,R., Tsukiyama,T., Ossipow,V. and Wu,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-AUG-1998) LNCB, National Cancer Institute, NIH,		
FEATURES	Building 37, Room 5E-20, Bethesda, MD 20892-4255, USA		
Source	Location/Qualifiers		
	1..873	/Organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"		
	/chromosome="2"		
	/map="60C"		
	1..873	/gene="Nurf-38"	
	1..873	/note="nucleosome remodeling factor complex component"	
	/codon_start=1		
	/product="inorganic pyrophosphatase Nurf-38"		
	/protein_id="AAC97112.1"		
	/db_xref="PID:g3600094"		
	/db_xref="GI:3600094"		
	/translation="MALTYEVVEGAKNSPSLYFKKKKCNVISPMDIFLYANEKKT IYNNVVEVPRTNAKMETSUKTPMNPDKODIKKGKLEFVANCFEPHKGYINNYSALPOT WENPDHETSTGCKGDNDPLDIVLEIGYRAVRKRDVLKVGLGPFALIDESETPWKIIA IDNVDPASKRVNDIADVDQYFGPLLRLATVMFKIYRKIPDGCKPENOFAPNGDANKADPA NTIIAEHFWMNLVHOSPASGISISTNITNIRNSEHIHPREAEKILAEKPNDSGVDEE VSQTVDMWHHLTK"		
BASE COUNT	228 a	247 c	233 g 165 t
ORIGIN			
Query Match	11.1%	Score 142:	DB 22: Length 873:
Best Local Similarity	63.8%	Pred. No. 1.44e-88;	
Matches 329; Conservative	0;	Mismatches 187;	Indels 0; Gaps 0;
Db	140	ACATGTTGGTGGAGGTCGCTTCACCGCTTTCACATACTGCTTCACATTTCCATTCACACA	451







```

repeat_region /note="Alusg repeat: matches 1. .299 of consensus"
4986. .5106
repeat_region /note="Charlie2 repeat: matches 34. .149 of consensus"
5266. .5553
repeat_region /note="Alusx repeat: matches 1. .312 of consensus"
5554. .5859
repeat_region /note="Charlie2 repeat: matches 2067. .2378 of consensus"
5860. .6145
repeat_region /note="Alusx repeat: matches 3. .290 of consensus"
6146. .7010
repeat_region /note="Charlie2 repeat: matches 2378. .3214 of consensus"
7441. .7543
repeat_region /note="L2 repeat: matches 2609. .2710 of consensus"
8063. .8176
repeat_region /note="57 copies 2 mer tt 65% conserved"
8239. .8547
repeat_region /note="Aluya8 repeat: matches 1. .308 of consensus"
8932. .9225
repeat_region /note="Alusg repeat: matches 1. .294 of consensus"
complement(8227. .9569)
misc_feature /gene="AJ1042K10.4"
9964. .10284
repeat_region /note="match: GSS AQ170974"
12826. .12925
repeat_region /note="L2 repeat: matches 2062. .2388 of consensus"
13772. .14070
repeat_region /note="MIR repeat: matches 33. .134 of consensus"
15998. .16220
repeat_region /note="Alusg repeat: matches 1. .298 of consensus"
18301. .18603
repeat_region /note="MIR repeat: matches 51. .262 of consensus"
18627. .18860
repeat_region /note="Aluy repeat: matches 1. .303 of consensus"
19526. .19819
repeat_region /note="Aludc repeat: matches 16. .287 of consensus"
20494. .20787
repeat_region /note="Alusc repeat: matches 1. .294 of consensus"
21016. .21309
repeat_region /note="match: GSS AQ002626"
21447. .21511
repeat_region /note="Alusx repeat: matches 1. .293 of consensus"
21585. .21640
repeat_region /note="Alu repeat: matches 244. .298 of consensus"
21641. .21950
repeat_region /note="MER44A repeat: matches 192. .243 of consensus"
21951. .22119
repeat_region /note="Aluy repeat: matches 2. .310 of consensus"
22428. .22726
repeat_region /note="MER44A repeat: matches 2. .192 of consensus"
23734. .24027
repeat_region /note="Alusg repeat: matches 1. .299 of consensus"
24423. .24690
repeat_region /note="AluyB8 repeat: matches 1. .303 of consensus"
24876. .24915
repeat_region /note="LIMC1 repeat: matches 6053. .6324 of consensus"
25082. .25991
repeat_region /note="MER5A repeat: matches 119. .159 of consensus"
25992. .26035
repeat_region /note="match: ESTs AI006343 AI020127 AI011875 NS9014
AI25767 AI01188 AI787828 AA699725 AA648465 AA651607
C87885 AA893853 AA782624 N64803 AA924572 AA927683 AI241181
AA632509 AI075774 AA811649 NS8210 AA631607"
26107. .26176
repeat_region /note="22 copies 2 mer aa 82% conserved"
27265. .28268
repeat_region /note="35 copies 2 mer gt 69% conserved"
27265. .28268
repeat_region /note="match: EST N41161"
27265. .28268
repeat_region /note="match: ESTs T19776 AA636314 AA930547 H63439 W78868
AA234323 AA053842 AA464134 W84723 AI126636 AA262994

```

```

misc_feature H63387"
28689. .29182
prim_transcript /note="match: GSS B14190"
28982. .29740
prim_transcript /note="match: ESTs AA307016 AA375298 H73308 AA247913
H73348"
29853. .31247
repeat_region /note="match: ESTs AA058913 AA218638 AA053233 RA0376
AA214074 AI167714 M78455 AA68885 NS8052 AA218712 H73309"
31248. .31279
prim_transcript /note="16 copies 2 mer aa 91% conserved"
32552. .32944
misc_feature /note="match: EST AA136901"
33277. .33464
prim_transcript /note="match: STS G33258"
complement(33277. .33774)
prim_transcript /note="match: ESTs Z42319 AI060394 W87941 W87940"
complement(33809. .35251)
AA809782 AA521409 AA400214 AA400125 T92466 AA572017
AA704714 AI248843 AA773853 AA136902 AA724710 AA629354
W36698 AA287801 AA136590"

Note: remainder of annotations omitted.

Query Match 10.8% Score 138. DB 29; Length 184180;
Best Local Similarity 78.5%; Pred. No. 2,98e-85;
Matches 204; Conservative 0; Mismatches 54; Indels 2; Gaps 2;

Db 135720 AAAGCATGAGTGTGTATATACACAGCATTCACAGCCCT-CAAGTGTATCCTGAC 135778
|||||
QY 800 AAAGCATGAGTGTGTATATACACAGCATTCCTGAGAGCCCTTCAAGTGTATCCTGAC 859
|||||
Db 135779 AAAGTGAAGCCATTTGTGATGCTTCCACACACATGATACCTGCTATGATACACG 135838
|||||
QY 860 GCGCCAGAGCCATTTGTGATGCTTCCACACACCTGATGATGCTGCGAC-ATATAC 918
|||||
Db 135839 CACAGATGTCGACGCTGCTTCCACACAGAGAAAATGAATCTTCTTGAATAGC 135898
|||||
QY 919 AACGACGTCGATAGTGTGATGCTTCCACACAGAAAATGAATGATTTCTGGAATAC 978
|||||
Db 135899 AGCTGATATGCTGCTGCTATATCATCTGGAAGTATGATGCAAGTACAGCCTTTC 135958
|||||
QY 979 AGCTGATATGCTGATCATGCTGCTATGATGATGATGATGATGATGATGATGATGATG 1038
|||||
Db 135959 AAAGCTTAAGCTTTAGAA 135978
|||||
QY 1039 AAAGCTTAATTTGTAGAA 1058
|||||

RESULT 7
LOCUS SCPAG 1612 bp DNA PLN 12-SEP-1993
DEFINITION Yeast PPA gene for inorganic pyrophosphatase (EC 3.6.1.1).
ACCESSION X13253
NID 94198
VERSION X13253.1 GI:4198
KEYWORDS inorganic pyrophosphatase; phosphatase; PPA gene; pyrophosphate
phosphohydrolase.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1612)
AUTHORS Kolakowski, L.F.
TITLE Submitted (14-OCT-1988) Kolakowski L.F., University of
Pennsylvania, Department of Chemistry, 231 South 34th Street,
Philadelphia, PA 19104
2 (bases 1 to 1612)
AUTHORS Kolakowski, L.F. Jr., Schlosser, M. and Cooperman, B.S.
TITLE Cloning, molecular characterization and chromosome localization of
the inorganic pyrophosphatase (PPA) gene from S. cerevisiae
JOURNAL Nucleic Acids Res. 16 (22), 10441-10452 (1988)

```

```

MEDLINE 89083474
FEATURES
source
1..1612
/organism="Saccharomyces cerevisiae"
/strain="X2180"
/db_xref="taxon:4932"
/chromosome="chromosome 2"
/clone.lib="lambda gt10.1"
/clone="PM14"
1..1612
/gene="IPPI"
/db_xref="SGD:L0000872"
4..9
/promoter
/gene="IPPI"
/note="TATA-like box"
/db_xref="SGD:L0000872"
69..73
/misc_feature
/gene="IPPI"
/note="pot. transcription start site"
/db_xref="SGD:L0000872"
109..117
/promoter
/gene="IPPI"
/note="TATA-like box"
/db_xref="SGD:L0000872"
139..143
/misc_feature
/gene="IPPI"
/note="pot. alt. transcription start site"
/db_xref="SGD:L0000872"
216..220
/misc_feature
/gene="IPPI"
/note="pot. alt. transcription start site"
/db_xref="SGD:L0000872"
227
/misc_feature
/gene="IPPI"
/note="pot. alt. transcription start site"
/db_xref="SGD:L0000872"
233..249
/promoter
/gene="IPPI"
/note="constitutive promoter"
/db_xref="SGD:L0000872"
253..260
/promoter
/gene="IPPI"
/note="TATA-like box"
/db_xref="SGD:L0000872"
291..297
/promoter
/gene="IPPI"
/note="TATA-like box"
/db_xref="SGD:L0000872"
361
/misc_feature
/gene="IPPI"
/note="pot. alt. transcription start site"
/db_xref="SGD:L0000872"
390..1253
/CDS
/gene="IPPI"
/note="pyrophosphatase (AA 1 - 287)"
/codon_start=1
/db_xref="SGD:L0000872"
/protein_id="CAA31629.1"
/db_xref="PIR:g4199"
/db_xref="GI:4199"
/translaton="MTYTRIGAKNTLEKVIYIKQKRVSAFHDIPIYADKENIT
NMVVEPRMTNAKLEIKETLNPIDIDTQKGLRVRNCFPHGITHNAGPQWE
DNPVSHETKAYGNDNPIDLEIGETIATYGVOKVKAUGIMALDDEGETDMVAID
INPLPAKLDNDIEVEKTFPGGLRAINEMFRITKIDGKRNQFAFSGEAKKKKVALD
IIEKTHSMQOLIAKSSDSKGIIDLNVILDPDPTVSKAASDAIPPAISKADAPIDS
IDKWFETSGSV"
BASE COUNT 496 a 321 c 306 g 489 t
ORIGIN
Query Match 10.3% Score 131; DB 27; Length 1612;
Best Local Similarity 61.8%; Pred. No. 1,82e-79;
Matches 344; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

```

```

Db 533 ACATGTTGTTGAATTCACGTTGGACCAAGCCAACTAGAAATCCAGGAAGAA 582
Oy 222 ACATGTTGTTGAATTCACGTTGGACCAAGCCAACTAGAAATCCAGGAAGAA 281
Db 563 CTTTAACCCAAATCATCCAAACACCAAGAGGCAAGTGAATTTGTTGAAGATGTT 642
Oy 282 CTTTAACCCAAATCATCCAAACACCAAGAGGCAAGTGAATTTGTTGAAGATGTT 341
Db 643 TCCCTCACCATTGTTACATTCACAACTATGTTGTTTCCCAAACTTGGAAACCCAA 702
Oy 342 TCCCTCACCATTGTTACATTCACAACTATGTTGTTTCCCAAACTTGGAAACCCAA 401
Db 703 AGCTAAGCCACCCAGAAACTAAGCAGTGTGACAAACAGATCCAAATGATGTGGAAA 762
Oy 402 GGCACAAATGATTAACATACATGCTGTGTGTGTGACAAATGACCAATGATGTGTA 461
Db 763 TTGTTGAATCTATGCTTACATGTCGTCAGTCAAGCAGGTTAAGGCTCTAGATATC 822
Oy 462 TTGTTGAATCTATGCTTACATGTCGTCAGTCAAGCAGGTTAAGGCTCTAGATATC 521
Db 823 CTTTATGATGAGAGTGAGAGCAGATTCGAAAGTATTCGATGATATTAACATCCAT 882
Oy 522 CTTTATGATGAGAGTGAGAGCAGATTCGAAAGTATTCGATGATATTAACATCCAT 581
Db 883 TAGCCCAAAATTAAGCAGATTCGAAAGTATTCGAAAGTATTCGAAAGTATTCGAA 942
Oy 582 ATGAGCAGCAATTAATGATGATTCATGATGATGATGATGATGATGATGATGATG 641
Db 943 CTACTAAGCAATGCTTCAATTTTCAAAATCCAGATGTTAAGCCAGAAACCAATTTG 1002
Oy 642 CTACTAAGCAATGCTTCAATTTTCAAAATCCAGATGTTAAGCCAGAAACCAATTTG 701
Db 1003 CCTCTCCGCTGAGAGTCAAGCAAGAGTATTCGATGATGATGATGATGATGATG 1062
Oy 702 CGTTTATGAGAGTATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 761
Db 1063 ACTCTCGAAGCAATTA 1079
Oy 762 ACCATGGAAGCAATTA 778

```

```

RESULT 8
LOCUS SCYBR011C 2411 bp DNA PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome II sequencing frame ORF YBR011C.
ACCESSION Z35880 Y13134
NID 9536205
VERSION Z35880.1 GI:536205
KEYWORDS
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
1 (bases 1 to 2411)
Entlan,K.D., Koetter,P., Rose,M., Li,Z., Thermann,R., Brendel,M.,
Baur,A., Boles,E., Miosga,T., Schaaff-Gerstenschlaeger,I. and
Zimmermann,F.K.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2411)
MIPS
JOURNAL Direct Submission
TITLE Submitted (30-AUG-1994) Data collected by MIPS on behalf of the
European yeast chromosome II sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
3 (bases 1 to 2411)
Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Billeau,N., Boles,E., Brandt,T.,
Brendel,M., Bruckner,M., Bussereau,F., Christiansen,C.,
Contreras,R., Crouzet,M., Cziepluch,C., Demolis,N., Delaveau,T.,
Doignon,F., Domdey,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakoury,M., Entlan,K.D., Feuerhahn,M., Fiers,W., Fobio,G.M.,
Fritz,C., Gassenhuber,H., Glansdorff,N., Goffeau,A., Grivell,L.A.,

```

de Haan, M., Hein, C., Herbert, C. J., Hollenberg, C. P., Holmstrom, K., Jacq, C., Jacquet, M., Jauniaux, J. C., Jonniaux, J. L., Kallio, T., Klesau, P., Kirchbach, L., Koetter, P., Korol, S., Lieb, S., Logghe, M., Lohan, A. J. E., Louis, E. J., Li, Z. Y., Maat, M. J., Mallet, L., Mannhaupt, G., Messenguy, F., Miosga, T., Molemans, F., Mueller, S., Nasir, F., Obermayer, B., Perea, J., Pierard, A., Pirivandi, E., Pohl, F. M., Pohl, T. M., Potler, S., Proft, M., Punelle, B., Ramezani, Rad, M., Rieger, M., Rose, M., Schaff, Gerstenschlaeger, I., Scherens, B., Schwarzlose, C., Skala, J., Slonimski, P. P., Smits, P. H. M., Souciet, J. L., Steensma, H. Y., Stucka, R., Uristatza, A., van der Aart, O. J., van Dyck, L., Vassarotti, A., Vetter, I., Vierendeels, F., Vissers, S., Wagner, G., de Wergifosse, P., Wolfe, K. H., Zagalski, M., Zimmermann, F. K., Mewes, H. W. and Klein, K. Complete DNA sequence of yeast chromosome II

EMBO J. 13 (24), 5795-5805 (1994)  
95112788

TITLE  
JOURNAL  
MEDLINE  
FEATURES  
2 source  
Location/Qualifiers  
1. 2411  
/organism="Saccharomyces cerevisiae"  
/db\_xref="taxon:4932"  
/chromosome="II"  
complement(51..61)  
/note="ARS-consensus"  
complement(374..1237)  
/gene="IP1"  
/db\_xref="SGD:L0000872"  
complement(374..1237)  
/gene="IP1"  
/note="ORF YBR011c"  
/codon\_start=1  
/db\_xref="SGD:L0000872"  
/protein\_id="CAA8494.1"  
/db\_xref="PID:g536206"  
/db\_xref="GI:536206"  
/db\_xref="SWISS-PROT:P00817"  
/translation="MTYTRQIGAKNTLEKVKYIENQKPVSAFDIPLADKENNIF  
NMVEIPRMNTAKLEIKETLINIOTDKTRPKNCPHNGYHNGARPMOTE  
DPVSHETKANGDNIDPIDLEIGETIATYQKQVAYALIMALDGETDWMYIAD  
INDPLAKNDIEDVEKIPGLKALNEMERIKIPGKPNQAFSGEKKNNKYAD  
IKETDMSKQLANGKSSDKGIDLNTVLTLPDTPYKASDAIPASPADAPADIDS  
IDWFFISGVS"  
complement(1932..2275)  
/note="delta remnant"  
BASE COUNT 733 a 435 c 445 g 798 t  
ORIGIN  
Query Match 10.3% Score 131; DB 27; Length 2411;  
Best Local Similarity 61.8%; Pred. No. 1,82e-79;  
Matches 344; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

Db 908 TTCGGTGGCTTACGTTGGCTTCCCAAGTTTGGGGAACCACTAGTGTGAAT 967  
Cp 418 ATGTTATCATATGGTCCCTGGCTTCCCAAGTGTGAGGATGACCACTTCAGAT 359  
Db 968 GTACCAATGCGGAGGAACAGTTTCAACAATCTCAACTGGCTTGTGTGTG 1027  
Cp 358 ATATCTTTATACGGGAACAATTCGCAATACGAGATTTCTTTTACATCTGT 299  
Db 1028 GATGATGGGTCAAGTTTCTTCCCTGGTATTTCAACTTGGCGTGTCCACGTG 1087  
Cp 298 TTTATAGGGTTTAAGGTCCTTTAGCAATCTCATTTTGCATTAGACCAACGCGTG 239  
Db 1088 AATTCAACAACCATGT 1104  
Cp 238 TACTTCAACTACCATGT 222

RESULT 9  
LOCUS PPRP1877 2326 bp DNA PLN 11-AUG-1998  
DEFINITION Pichia pastoris TRP1 and IP1 genes.  
ACCESSION AF001000  
NID 92624377  
VERSION AF001000.1 GI:2624377  
KEYWORDS inorganic pyrophosphatase; IP1 gene;  
N(5'-phosphoribosyl)anthranilate isomerase; pyrophosphate  
phosphohydrolase; TRP1 gene; tryptophan biosynthesis.  
Pichia pastoris.  
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
Saccharomycetaceae; Pichia.  
1 (bases 1 to 2326)  
Molina, M.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (05-AUG-1997) Molina M., Departamento de Microbiologia  
JOURNAL II, Universidad Complutense de Madrid, Plaza de Ramon y Cajal s/n,  
E-28040 Madrid, Spain  
REMARK  
AUTHORS 2 (bases 1 to 2326)  
Cosano, I., Alvarez, P., Molina, M. and Nombela, C.  
TITLE Cloning and sequence analysis of the Pichia pastoris TRP1, IP1 and  
HIS3 genes  
Yeast 14 (9), 861-867 (1998)  
99034033  
3 (bases 1 to 2326)  
Molina, M.  
REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (13-NOV-1997) Molina M., Departamento de Microbiologia  
JOURNAL II, Universidad Complutense de Madrid, Plaza de Ramon y Cajal s/n,  
E-28040 Madrid, Spain  
FEATURES  
source  
1. 2326  
/organism="Pichia pastoris"  
/plasmid="pRS316-GAL1"  
/strain="NRRL Y-11430"  
/db\_xref="taxon:4922"  
350..1063  
/gene="TRP1"  
350..1063  
/EC-number="5.3.1.24"  
/function="third step in tryptophan biosynthesis pathway"  
/codon\_start=1  
/product="N(5'-phosphoribosyl)anthranilate isomerase"  
/protein\_id="CAA04452.1"  
/db\_xref="PID:ell80017"  
/db\_xref="PID:g2624378"  
/db\_xref="GI:2624378"  
/db\_xref="SWISS-PROT:O13504"  
/translation="MALVKICGLQSLAEATVANGASLVGVINVPGERTVQEVAR  
EISOMVERKISGSKYLDSDROLDKREWDCEPLEDMFENYKKEISSGPIVGEVRNOS  
IDELCOAIIHGHDFVQHGSPEDSYIRNIPVITRYDNDNAVDCITGEDIINRAL  
VLDSDEGGGSKTIDMARAKFGERRKYLILAGVYTPDVAHAHRSHTCIGVDVSGCV

gene  
CDS  
661 a 488 c 540 g 636 t 1 others

ETNAASKDMDKITQCFIRNAT"  
complement(1110, .1967)  
/gene="lppi"  
complement(1110, .1967)  
/gene="lppi"  
/EC\_number="3.6.1.1"  
/codon\_start=1  
/product="inorganic pyrophosphatase (pyrophosphate  
phospho-hydrolase)"  
/protein\_id="CA040453.1"  
/db\_xref="PID:e1180018"  
/db\_xref="PID:g2624379"  
/db\_xref="GI:2624379"  
/db\_xref="SWISS-PROT:O13505"  
/translation="MSYSTROIQGANLTLENVFLKDGQVSPPHDIPILADESKYVL  
NMVVEYPRMTNAKLEISKELNPLIDDTTKGKRLFRPNCCPHHGYTHNTGAFQIWE  
DPNATHEETAKGNDPLDVCEIGERSTYGGQKVKLVGVALLLDEGETDKRVYIDID  
NDPLAPPLNIDIEYEHMPGILLRATNEMFRYIKLPDCKPENQAFSECKKKYAEVEV  
IOEERAMEKILIEFKTSPEGLDNTNTLTQSPSPSATSVPYASPAAPAKIDQSIDJ  
KMFYSSSA"

[illegible]

ORGANISM	Schizosaccharomyces pombe
Eukaryota:	Fungi: Ascomycota: Archiascomycetes:
Schizosaccharomycetales:	Schizosaccharomycetaceae:
Schizosaccharomyces:	
REFERENCE	1 (bases 1 to 1366).
AUTHORS	Kawasaki, I.
TITLE	Direct Submision
JOURNAL	Submitted (08-MUG-1990) Kawasaki I., The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai Minato ku Tokyo 108, Japan
REFERENCE	2 (bases 1 to 1366)
AUTHORS	Kawasaki, I., Adachi, N. and Ikeda, H.
TITLE	Nucleotide sequence of S. pombe Inorganic pyrophosphatase
JOURNAL	Nucleic Acids Res. 18 (1'9), 5888 (1990)
MEDLINE	91016938
COMMENT	Data kindly reviewed (06-NOV-1990) by Kawasaki I.
FEATURES	Location/Qualifiers
source	1 - 1366

[illegible]





Db 7422 ATGCTGAGAGACCATCTTGAACATGGTTCGTTGAGATCCCTCGTTGGACTCAGGCTAAGC 7481

QY	201	ATGCGAGTATGAGGATGCTGTTTCACATGAGTAGTGAAGTACCACGCTGGCTTAATGCANAAA	260
Db	7482	TGGAATTAACCAAGGAGACTACCTGAAACCCCTATACAGCAAGATTACTAAGAGGTAGC	7541
QY	261	TGGAGATTGCTTAACAAAGGAGACCCCTTTAAACCCCTATTAAACAAAGATGTAAAAAAGGAAAC	320
Db	7542	TTTCGTTGCTGTAACAGCTTTCTCCACACGCGTTACATTGGAGCTACGGCTTTTC	7601
QY	321	TTCCGATATGTTGCCGAATTTGTTCCCGATATAAAGATATATCTGAACTATATGTCATCC	380
Db	7602	CTCAACCTATAGAGATCTCTATGTTTCACTCTCTGAACCAAGGCCAAGGCTGATAGCG	7661
QY	381	CTCAACTTGGGAAGACCCAGGCAACATGATTAACAATCTGCTGCTTGTGGTACATG	440
Db	7662	ATCCCTCTGATGTTTGTGAGATCGTGAGCTGTGCTACACCGCCAGTCACGACG	7721
QY	441	ACCCAAATGATGTGTGTAATTTGGAAGCAAGATATGTGCAAGAGGGAATTAATTTGGC	500
Db	7722	TTAAGTCTTTGGGCTTAATGACCCCTTTCTTACAGAGGCTGAGCTACTGAGAGGTTATG	7781
QY	501	TGAAGTCTTAGGGAATATGCTTAATGATTAACGAAGGCGAAACCGACTGAAAGCTATTG	560
Db	7782	TCATGAGCTTAATGAGCCCTTTACTCTTACCTTAAGCTTAATGATTTGAGAGTGAAGCTG	7841
QY	561	CCATTAAATGATGATGATGATGCTGATGACCACTTAATTAATGATTAATGATTAATGATTA	620
Db	7842	ATATGCTGCTTAATGCTGCTACTACTACGAGTGTGCTCCATTTACAGATCCCGATG	7901
QY	621	TGAACCTGCTGCTACTACTACTGAGCTACTGTGAGCTGCTTAAGAGTAAAGGTTCCGATG	680
Db	7902	GCAAGCCGAAATTTCTCTTCTGCTTCTGATGCTGAT	7937
QY	681	GAAACCAAGAAATGACTTTCGCTTAATGACGAAT	716
RESULT	12.		
LOCUS	AF085600	2826 bp	DNA
DEFINITION	Drosophila melanogaster	inorganic pyrophosphatase	NURF-38 (Nurf-38)
ACCESSION	AF085600	gene complete cds.	
NID	G3746879		
VERSION	AF085600.1	GI:3746879	
KEYWORDS			
SOURCE			
ORGANISM			
	fruit fly.		
	Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Diptera; Brachycera; Muscomorpha; Pphyroidea;		
	Drosophilidae; Drosophila.		
	1. (bases 1 to 2826)		
REFERENCE	Guala, D.A., Sandatizopoulos, R., Tsukiyama, T., Ossipow, V. and Wu, C.		
AUTHORS	Inorganic pyrophosphatase is a component of the Drosophila		
TITLE	nucleosome remodeling factor complex		
JOURNAL	Genes Dev. 12 (20), 3206-3216 (1998)		
MEDLINE	9903073		
REFERENCE	2 (bases 1 to 2826)		
AUTHORS	Guala, D.A., Sandatizopoulos, R., Tsukiyama, T., Ossipow, V. and Wu, C.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-AUG-1998) LNCB, National Cancer Institute, NIH,		
	Building 37 Room 5E-20, Bethesda, MD 20892-4255, USA		
FEATURES			
source	1..2826		
	/organism="Drosophila melanogaster"		
	/db_xref="taxon:7227"		
	/chromosome="2"		
	/map="60C"		
mRNA	join(556..704,1148..1267,1333..1728,1784..1935,1994..2102,		
	2422..2585)		
	/gene="Nurf-38"		
	/product="NURF-38"		
gene	554..2585		
	/gene="Nurf-38"		
CDS	join(642..704,1148..1267,1333..1728,1784..1935,1994..2102,		



```

source
1. 37013
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/clone="cosmid c3A12"
/map="IL"
/complement(1. 1114)
/gene="SPAC3A12.01"
1. 110
/gene="SPAC3A12.01"
/note="nominal overlap with c20G8"
/complement(1. 1114)
/partial
/gene="SPAC3A12.01"
/note="SPAC3A12.01c, similarity: to YPL10W, Q02948,
LPH/P, (557aa), fasta scores, opt:391, E():4.9e-16, (25.68
identity in 344 aa overlap)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB08758.1"
/db_xref="PID:e351362"
/db_xref="PID:g2462674"
/db_xref="GI:2462674"
/db_xref="SWISS-PROT:P87117"
/translation="MOYLCOCHSLINKKDYDDDLKIKLPLKSRFYQASLTENME
SGESDDDNSSSEDPAGRLQLYKKTISEGYNNDNDNPPRLPTLDSFVPLPAAD
GEEKNSPEEVNLFSSKILTYNRIPDLSSKRVVHPPLCYEACALLTEMSKTLA
LKEKKMYNDNLSSOTVEENPAIDSLIDELMKQINKKEKIEISDETQKQ
LRLDEKEVVAEDEFNNNOFOFKTSLEROYDCANLEFEHNSRKLKJOKKN
VFSDIYSHSEPNESGIATINGLRGRGPSOKVMVAEINAMGMVILLDLVLEK
LDFSSSYOLKPFSGSFIIRFDRDPNGNQYKPKLD"
1796. 1868
/note="tRNA-Cys, anticodon gca, length = 73"
2563. 3589
/gene="SPAC3A12.02"
/join(2563. 2718, 2827. 3199, 3258. 3589)
/gene="SPAC3A12.02"
/note="SPAC3A12.02, len:286aa, similar eg. to IPYR_SCHPO,
P19117, inorganic pyrophosphatase, (286 aa), fasta scores,
opt: 930, E():0, (48.7% identity in 277 aa overlap), also
similar eg. to YBR011C, IPYR_YEAST, P00817, inorganic
pyrophosphatase, (286 aa), fasta scores, opt: 904, E():0,
(54.4% identity in 250 aa overlap)"
/codon_start=1
/product="inorganic pyrophosphatase"
/protein_id="CAB08747.1"
/db_xref="PID:e316097"
/db_xref="PID:g2104418"
/db_xref="GI:2104418"
/db_xref="SWISS-PROT:P87118"
/translation="MASLAKNIILOFRSKIKINTPPRYCYKNNKPISEFHDVPLT
SDKDTMVMWEIPEWPAKCEISLTSFPHIKDILKNGKLRVANSFPIGFHWNGA
LPQWEPPNVIDSKTKMGGDPLDVEICIGSISGIIQVAKVAGLGLIDGEGTDG
KILAIIDNPRAKLNDISDVONLMPRLCTCRWFPIIYIIPGCKPNRFPGNYLP
KSAALDIACOHOMKYSRDKOYIKNFHNSVNNVDLINKINSIKEEVSQNSVNS
FPHYHILPNL"
2713. 3267
/gene="SPAC3A12.02"
/note="pfam match to entry PF007719 Pyrophosphatase,
Inorganic pyrophosphatase"
2719. 2724
/gene="SPAC3A12.02"
/note="gtagct, splice donor sequence"
2806. 2826
/gene="SPAC3A12.02"
/note="ctaacatttttcaactag, splice branch and acceptor
sequence"
3034. 3054
/gene="SPAC3A12.02"
/note="PS00387 Inorganic pyrophosphatase signature"
3200. 3205
/gene="SPAC3A12.02"






misc_feature
3245. 3257
/note="gtagct, splice donor sequence"
/gene="SPAC3A12.02"
/note="ctaacattcttag, splice branch and acceptor sequence"
complement(3887. 4816)
/gene="SPAC3A12.03c"
complement(3887. 4816)
/gene="SPAC3A12.03c"
len:309aa, similar eg. to g_gallus
090972, RING ZINC FINGER PROTEIN, (381 aa), fasta scores,
opt:214, E():1.4e-07, (37.3% identity in 67 aa overlap)
also similar to GOL1_DROME, Q06003, gollath protein (91
protein), (284 aa), fasta scores, opt:203, E():6e-07,
(36.5% identity in 74 aa overlap)"
/codon_start=1
/product="zinc finger, C3HC4 type"
/protein_id="CAB08748.1"
/db_xref="PID:e315874"
/db_xref="PID:g2104419"
/db_xref="GI:2104419"
/db_xref="SPTRMBL:P87119"
/translation="MPIPKNRPMPHVEEVSQNTTEILLFALYIISVIRINFEFYL
RCQVYFHTLENOGDDEPRLIOHNVNRSSTGLSPVRLNGLVGYDIPSRRSV
VSKRALSCISLEIPIYIKMLKKRGHAKGESETPVNSQSVYVOCGEPSSYITTD
VRPNIGSTSEFVEMSSLSNITYNTDASDDSDSCLEDEDFCTIADYAFDIL
RVLPCEBVFHTQCIDTWITTKASCPLCNDYIKYFLQMDAASVTHENAMSIPLSP
GDSRTSHAEIDRSLLSMSVSRNSRMPYIVYSTL"
complement(4079. 4204)
/gene="SPAC3A12.03c"
/note="pfam match to entry PF00097 zf-C3HC4, zinc finger,
C3HC4 type"
complement(5344. 6105)
/gene="SPAC3A12.04c"
complement(join(5344. 5808, 5878. 5976, 6019. 6105))
/gene="SPAC3A12.04c"
/note="SPAC3A12.04c, len:216aa, similar eg. to YHR062C,
YHM2_YEAST, P38786, hypothetical 32.2 kd protein, (233aa),
fasta scores, opt:420, E():4e-22, (34.9% identity in 218
aa overlap)"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAB08749.1"
/db_xref="PID:e316098"
/db_xref="PID:g2104420"
/db_xref="GI:2104420"
/db_xref="SWISS-PROT:P87120"
/translation="MINFYLOKVGATALANVOYDGKLONYKINPIKVEIKVEKIK
IYSRILITLESMPONKVLNVTKEPILAIRPGDRLLOOTCSDEFDLISDFQRL
PFYLRKTFMGIAVSRDGIETISGSRVNSRNRLITNATISVYRATRGRTIIVSET
KPILECRACFDVYNLATFMDLKODOKARKSVGESCNSVLLHAETRRDTYSILNGCH"
complement(5809. 5823)
/gene="SPAC3A12.04c"
/note="ctaacatttccttag, splice branch and acceptor
sequence"
complement(5872. 5877)
/gene="SPAC3A12.04c"
/note="gtagct, splice donor sequence"
complement(5977. 5985)
/gene="SPAC3A12.04c"
/note="ctaacattag, splice branch and acceptor sequence"
complement(6013. 6018)
/gene="SPAC3A12.04c"
/note="gtagctg, splice donor sequence"
complement(6860. 10301)
/gene="SPAC3A12.05c"
complement(join(6860. 7795, 7836. 10215, 10291. 10301))
/gene="SPAC3A12.05c"
/note="SPAC3A12.05c, putative transcription initiation
factor, len:1108aa, similar eg. to YCR42C, T2D2_YEAST,
P23253, transcription initiation factor tsm1, (1407aa),
fasta scores, opt:1598, E():0, (31.0% identity in 1226 aa
overlap)"
/codon_start=1

```







Release 3.1A John F. Collins, Biocomputing Research Unit  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

```

Mpsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Dec 27 12:53:55 1999; MasPar time 277.57 Seconds
Tabular output not generated. 984.229 Million cell updates/sec

```

```

Title: >US-09-415-540-2
Description: (1-1275) from US09415540.seq
Perfect Score: 1274
N.A. Sequence:
Comp: 1 CAAAGAGTTGGGCGTCTCT.....CAATTAATTAATTTGCTG 1275
      GTTCTCCANCCCGCAGAGA.....CTTATTTTAATAAACGAC

```

Scoring table: TABLE default

```
Nmatch      STD :  Dbase 0;  Query 0
```

Searched: 271905 seqs, 107135622 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database:

1:part2 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13 14:part14 15:part15 16:part16 17:part17 18:part18 19:part19 20:part20 21:part21 22:part22 23:part23 24:part24 25:part25 26:part26 27:part27 28:part28 29:part29 30:part30 31:part31 32:part32 33:part33 34:part34 35:part35 36:part36 37:part37 38:part38 39:part39 40:part40 41:part41 42:part42 43:part43 44:part44 45:part45 46:part46 47:part47 48:part48 49:part49 50:part50 51:part51 52:part52 53:part53 54:part54 55:part55 56:part56 57:part57 58:part58 59:part59 60:part60

Statistics: Mean 9.230; Variance 5.556; scale 1.661

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1274	100.0	1275	54	V63542	cDNA encoding a human	0.00e+00
2	1265	99.3	1351	60	X00685	Human secreted protein	0.00e+00
3	1258	98.7	1350	60	X00716	Human secreted protein	0.00e+00
4	330	30.6	416	169	T21836	Human gene signature	2.18e-24
5	44	3.5	91	9	O51746	Oligonucleotide probe	2.95e-11
6	42	3.3	204	1	N81164	Basse substituted E.co	4.25e-10
7	40	3.1	91	9	O51746	Oligonucleotide probe	5.93e-09
8	40	3.1	204	1	N81164	Basse substituted E.co	5.93e-09
9	36	2.8	114	12	O70465	Genetic DNA sequence	1.04e-06

C	10	36	2.8	114.12	070468	Generic DNA sequence	1.04e-06
C	11	36	2.8	169.32	T76307	Human RANFES antisens	1.04e-06
C	12	34	2.7	91.46	V44650	Mammalian DNA replica	1.29e-05
C	13	35	2.7	114.12	070469	Generic DNA sequence	3.68e-06
C	14	34	2.7	114.12	070467	Generic DNA sequence	1.29e-05
C	15	34	2.7	114.12	070467	Generic DNA sequence	1.29e-05
C	16	33	2.6	114.12	070468	Generic DNA sequence	4.47e-05
C	17	32	2.5	91.46	V44650	Mammalian DNA replica	1.53e-04
C	18	32	2.5	114.12	070470	Generic DNA sequence	1.53e-04
C	19	32	2.5	114.12	070469	Generic DNA sequence	1.53e-04
C	20	32	2.5	114.12	070465	Generic DNA sequence	1.53e-04
C	21	32	2.5	114.12	070466	Generic DNA sequence	1.53e-04
C	22	32	2.5	190.32	T76452	Chymase antisense oli	1.53e-04
C	23	31	2.4	114.12	070466	Generic DNA sequence	1.72e-03
C	24	30	2.4	114.12	070470	Generic DNA sequence	1.72e-03
C	25	30	2.4	114.12	070472	Generic DNA sequence	1.72e-03
C	26	31	2.4	117.32	T76363	Human Interleukin 8 a	5.17e-03
C	27	30	2.4	178.32	T76405	Human endothelin-1 an	1.72e-03
C	28	30	2.4	657.7	043519	Degenerate FMN reduct	1.72e-03
C	29	29	2.3	70.32	T76184	Human IL4 antisense o	5.66e-03
C	30	29	2.3	178.32	T76605	Human endothelin-1 an	5.66e-03
C	31	28	2.2	69.32	T76219	Human IL6 antisense o	1.83e-02
C	32	28	2.2	128.32	T76233	Human IL6 antisense o	1.83e-02
C	33	28	2.2	160.47	V48098	Oligonucleotide Lp160	1.83e-02
C	34	28	2.2	250.32	T76438	Substance P antisense	1.83e-02
C	35	27	2.1	88.32	T76170	Human IL3 receptor an	5.80e-02
C	36	27	2.1	114.12	070472	Generic DNA sequence	5.80e-02
C	37	27	2.1	114.12	070471	Generic DNA sequence	5.80e-02
C	38	27	2.1	130.47	V48104	Randomised Pool oligo	5.80e-02
C	39	27	2.1	160.47	V48098	Oligonucleotide Lp160	5.80e-02
C	40	26	2.0	114.12	070473	Generic DNA sequence	1.81e-01
C	41	26	2.0	140.47	V48104	Randomised Pool oligo	1.81e-01
C	42	26	2.0	136.21	T13624	ME#2b library generat	1.81e-01
C	43	25	2.0	264.32	T76445	Substance P receptor	5.53e-01
C	44	26	2.0	501.3	N50033	Sequence encoding new	1.81e-01
C	45	26	2.0	2539.51	V59609	Human secreted protei	1.81e-01

## ALIGNMENTS

	RESULT	1	
ID	V63542	standard; cDNA; 1275 BP.	
AD	V63542;		
DT	03-FEB-1999	(first entry)	
DE	cDNA encoding a human inorganic pyrophosphatase designated HYP.		
KW	Human; inorganic pyrophosphatase; HYP; Incyte clone 768330;		
OS	cell regeneration; tissue regeneration; ss.		
FT	Homo sapiens.		
FT	Key		
FT	Location/Qualifiers		
FT	CDS	89..958	
FT	/tag= a		
FT	/product= HYP		
PN	US5843665-A.		
PD	01-DEC-1998.		
PF	31-OCT-1996; 741437.		
PR	31-OCT-1996; US-741437.		
PA	(INCYTE) INCYTE PHARM INC;		
PI	Hawkins PR, Hillman JL;		
PI	WPI; 99-044570/04.		
DR	P-PDBD; W70876.		
PT	DNA encoding human pyrophosphatase protein - useful for producing recombinant protein, which may be useful in promoting cell and tissue regeneration		
PS	Claim 3; Fig 3A-D; 29pp; English.		
CC	The present sequence encodes a human inorganic pyrophosphatase designated HYP. Nucleic acids encoding HYP were first identified in Incyte clone 768330 from the lung tissue cDNA library LUNGNOTO4		
CC	The recombinant pyrophosphatase, may be useful in promoting cell and tissue regeneration.		
SC	Sequence 1275 BP; 394 A; 235 C; 294 G; 351 T;		
Query Match	100.0%; Score 1274; DB 54; Length 1275.		
Best Local Similarity	100.0%; Pred. No. 0.00e+00;		

Dd	1081.	tgcgtgactaatccacatatatcagaagtattaccataagcaatttttcatactcca	1140
Oy	1081.	tgcgttgactaatccacatatatcagaagtattaccataagcaatttttcatactcca	1140
Dd	1141	actaagaataacttttagcacatctttaataatcaaagcagttgttcatgttgaagcactt	1200
Oy	1141	actaagaataacttttagcacatctttaataatcaaagcagttgttcatgttgaagcactt	1200
Dd	1201	gtgaataagatgtcgaaaggagacaacattgtagtgtatattgttacccatagttaagaaat	1260
Oy	1201	gtgaataagatgtcgaaaggagacaacattgtagtgtatattgttacccatagttaagaaat	1260
Dd	1261	aaaattattttgctg	1275
Oy	1261	aaaattattttgctg	1275
<b>RESULT 2</b>			
ID	X00685	standard; DNA; 1351 BP.	
AC	X00685;		
DT	25-MAR-1999	(first entry)	
DE	Human secreted protein gene 75 clone H0SEI145.		
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;		
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;		
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;		
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;		
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;		
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclasts; thymus;		
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;		
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.		
OS	Homo sapiens.		
PN	W09842738-A1.		
PD	01-OCT-1998.		
PF	19-MAR-1998;	U05311.	
PR	30-MAY-1997;	US-050937.	
PR	21-MAR-1997;	US-041276.	
PR	21-MAR-1997;	US-041277.	
PR	21-MAR-1997;	US-041281.	
PR	21-MAR-1997;	US-042344.	
PR	30-MAY-1997;	US-048069.	
PR	30-MAY-1997;	US-048094.	
PR	30-MAY-1997;	US-048095.	
PR	30-MAY-1997;	US-048096.	
PR	30-MAY-1997;	US-048098.	
PR	30-MAY-1997;	US-048131.	
PR	30-MAY-1997;	US-048135.	
PR	30-MAY-1997;	US-048154.	
PR	30-MAY-1997;	US-048160.	
PR	30-MAY-1997;	US-048186.	
PR	30-MAY-1997;	US-048187.	
PR	30-MAY-1997;	US-048188.	
PR	30-MAY-1997;	US-048350.	
PR	30-MAY-1997;	US-048351.	
PR	30-MAY-1997;	US-048352.	
PR	30-MAY-1997;	US-048355.	
PR	05-AUG-1997;	US-054804.	
PA	(HOMA-) HUMAN GENOME.SCI INC.		
P1	Brewer LA, Duan R, Ebnert R, Ferrle AM, Florence KA,		
P1	Greene JM, Hu JS, Lafleur DW, Moore PA, Ni J, Olsen HS,		
P1	Rosen CA, Ruben SM, Shi Y, Young P;		
DR	WPI: 99-070066/06.		
DR	P-PDSB; W67881.		
PT	New isolated human genes and the secreted polypeptides they encode -		
PT	useful for diagnosis and treatment of e.g. cancers, neurological		
PT	disorders, immune diseases, inflammation or blood disorders		
PS	Claim 1: Page 233-234; 385pp; English.		
CC	This sequence represents a nucleic acid molecule which encodes a secreted		
CC	human protein. The gene number, and the clone it is derived from, are		
CC	detailed in the descriptor line. The gene can be used to generate fusion		
CC	proteins by linking to the gene to a human immunoglobulin Fc portion		
CC	(e.g. X00602) for increasing the stability of the fused protein as		
CC	compared to the human protein only.		



Pt	Rosen CA, Ruben SM, Shi Y, Young P:
Df	WPI: 99-070066/06.
Dr	P-PSDB; W67912.
Pt	New isolated human genes and the secreted polypeptides they encode
Pt	diseal for diagnosis and treatment of e.g. cancers, neurological
Pt	disorders, immune diseases, inflammation or blood disorders
Pt	Claim 1; Page 264-265; 385pp: English.
Cc	This sequence represents a nucleic acid molecule which encodes a secreted
Cc	human protein. The gene number, and the clone it is derived from, are
Cc	detailed in the descriptor line. The gene can be used to generate fusion
Cc	proteins by linking to the gene to a human immunoglobulin Fc portion
Cc	(e.g. X00602) for increasing the stability of the fused protein as
Cc	compared to the human protein only.
Cc	The invention relates to 87 novel genes and their fragments (nucleic acid
Cc	sequences: X00611-X00724, amino acid sequences W67807-W68004) which
Cc	are useful for preventing, treating or ameliorating medical conditions
Cc	e.g. by protein or gene therapy. Also, pathological conditions can be
Cc	diagnosed by determining the amount of the new polypeptides in a sample
Cc	or by determining the presence of mutations in the new polynucleotides.
Cc	Specific uses are described for each of the 87 polynucleotides, based on
Cc	which tissues they are most highly expressed in (see X00611 for described
Cc	uses).
SQ	Sequence 1350 BP: 414 A; 250 C; 319 G; 363 T;
Query Match	98.7%; Score 1258; DB 60; Length 1350;
Best Local Similarity	99.9%; Pred. No. 0.00e+00;
Matches 1264; Conservative	0; Mismatches 0; Indels 1; Gaps 1;
Df	20 ggggctctctctttgttcagtcgcgcgcgctcgccgctgtgtgtctcttggcaagcgcg 79
Oy	
Oy	11 GGGGCTCTCTCCTTTGCATGTAGTCGGCGCGGCTGGTGGCTCTGTGGAGCGCGG 70
Df	80 cggcagaactccggcacatgatagcggcttcacgaccagacggagcgcgcgcgc -ttctcc 138
Oy	
Oy	71 CGCGAGACACTCCGGCACTATGACGGCTTCAGACCAGAGACGCCGCGCCCTTTC 130
Df	139 ctggagracccgaattcttcctcaaaatggaaaaggacaatatattctccattcatgat 198
Oy	
Oy	131 CTGGAGTACCGGAGCTTCTCCCAAATAATGAAAGGACAATATATTCCTCATTCATGAT 190
Df	199 attccaattatgcagataagagtgtgtttccaatgtatgttggaagtcaccaagtcgct 258
Oy	
Oy	191 ATTCCATTATTATGCAGATAGATGTGTTTCACATGTGTAAGTAGTACACGCTGGCT 250
Df	259 aatgcacaaatlgagatltgctacaagaagcccttaaaccttataaacaagatgtgaa 318
Oy	
Oy	251 AATGCAAAAATGAGAATTGCTACAAAGGACCCTTAAACCCATTAAACAGATGTGAAA 310
Df	319 aaagaaaacttgcctatgttgtgaattgtgtccgtataaagatatatctcgaaacct 378
Oy	
Oy	311 AAAGAAAACCTTGCTATGTGTGGAAATTTGTCCTCCGTATAAAGATATATCTGGAACAT 370
Df	379 gtgtccatcccctcagacttggagaagccagagccacaatgataaacatctatgcgttgt 438
Oy	
Oy	371 GTGTCCATCCCCTCAGACTTGGGAAGCCAGGCGCAATGATAAACATATCTGCGTGT 430
Df	439 ggtgacaatgacccaattgtatgtgttgaatttggaaagaagatgtgtgcaagagtgtga 498
Oy	
Oy	431 GGTCACATATACCCAATTGATGTGTGAAATTTGGAACCAAGCATGTGCAAGAGTGAA 490
Df	499 ataattggttgaagaagtcttaggcctatgtggcctatgatgaagaagggaaccgcagctg 558
Oy	
Oy	491 ATATATTGGCTGGAAGTCTTAGGCATATTTGGCTATGATTAAGCAAGGGGGAACCCACAG 550
Df	559 aaagctatgccatlaattgtgatatctctatgcagcccaattcataatgataataatgat 618
Oy	
Oy	551 AAAGCTATTGCCATTATGTGGATGATCCTGATGACGCCAATTAATATGATATCAATAT 610
Df	619 gtcaaaagcgtgaacctgtgctacttagaagcctactgttgagctgtttagaagtgataag 678
Oy	
Oy	611 GTCAAAAGCGGTGAACCTGGCTACTTGAAGAGCTAGCTGAGCTTGAAGAGTATAG 670
Df	679 gtcttcgattggaanaaccagaataatgagtlitgcgtltaalcgcagaaattcaagaataagac 738

Qy	671	CTTCCGATGAGAAACCAAGAAATAGTTTCGCTTTATGCAATTTAAAGTAAGAC	73
Db	739	tttgcattgatataattataaaagccatctatgacatttggaaagcatagtgtactaa	79
Qy	731	TTTGCCATTGATATTATTAAAGCACTCATGACCATTTGAAAGATTAGTACTAAGAA	79
Db	799	acgaatggaaaagaagcaatcagttgtcatgtataacaaactttgtctgagaccctcaagt	85
Qy	791	ACGAATGGAAAAGCATCATCGTTGCATGTAATCAACTTTGCTGAGAGCCCCCTTCAAGTGT	85
Db	859	gattccgatgctgcccagagccatttggatgctttaccacacccctgtgattctgcttc	91
Qy	851	GATCCTGATGCTGCCAGACCACTTGTGGATCGTTTACACACCCTGTGATGTGCTTC	91
Db	919	acagttacccacagacagcttgatataagtggttccatccacagaaaaactaattpagatttctc	97
Qy	911	ACAGTACCAAAAGAGGTGATAGTGGTTCATGCATCCAGAAAACCTATGATGATTTCTCT	97
Db	979	ggaatacaagctgatatgtgtacatcgtgttccatctgtgattgataagtaagtagt	10
Qy	971	GGAATCAAGCTGATATTGCTTACATCCGTTTCATCTGGATGTGATTAAGATAAAGTACT	10
Db	1039	agcttttcaagctttaaattgtgtgaacctatcctaagaatlaatttcgtgtgact	10
Qy	1031	AGCTTTTCAAGCTTTAAATTTGTGAACTCATTTAACTAAGTAATTTGCTGTGACT	10
Db	1099	aatccaatactacagaaatgttatccatctaaagcaattttccatctcaactaagataa	11
Qy	1091	AATCCAATATACAGCAAGAAATGTTATTCATCTAAACATTTTTCATATCTCAACATAAATA	11
Db	1159	cttttgacacatgcttaaatcaatcaagaacagttgtcatttggaaagtcacttgyaataagat	12
Qy	1151	CTTTTGGACATGCTTAAATATCAAAACAGCTTTCATTTGGAAAGTCACTGTGTAATAGAT	12
Db	1219	gttcaaggaggagcaaatatgtgatatatttcacatgtagtaggaataaaataattatt	12
Qy	1211	GTCGAAGGGAGACATATTTGATGTATATTTTAACTAATGTATGGAAATAAATATTTT	12
Db	1279	tgtctt 1283	
Qy	1271	TGCTG 1275	
RESULT 4.			
ID	T21836 standard: cDNA to mRNA: 416 BP.		
AC	T21836;		
DT	01-AUG-1996 (first entry)		
DE	Human gene signature HMG6S3377.		
KW	Gene signature; messenger RNA; mRNA; relative abundance; frequency;		
KW	human; cloning; mapping; non-biased library; diagnosis; detection;		
KW	cell typing; abnormal cell function; ss.		
OS	Homo sapiens.		
PN	M09514772-11.		
PR	01-JUN-1995.		
EF	11-NOV-1994; J01916.		
PR	12-NOV-1993; JP-355504.		
PA	(MATS/) MATSUBARA K.		
PA	(OKUB/) OKUBO K.		
PI	Matsubara K, Okubo K;		
DR	WPI, 95-206931/27.		
PT	Identifying gene signatures in 3'-directed human cDNA library - e.g.		
PT	for diagnosis of abnormal cell function, by preparing cDNA that		
PT	reflects relative abundance of corresp. mRNA in specific human		
PT	tissues		
PS	Claim 1: Page 977; 2245pp; Japanese		
CC	A single-stranded DNA (or its complementary strand or the corresp.		
CC	double-stranded DNA) which comprises one of the 7837 "GS" sequences		
CC	given in 119001-r76837 and which is able to hybridise to part of		
CC	human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)		
CC	sequences were obtained from 3'-directed cDNA libraries prepared		
CC	from various human tissues; synthesis of cDNA was initiated from the		
CC	3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-		

CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types. 77 C: 128 T:  
50 Sequence 416 bp; 128 A; 77 G; 128 T;

Query Match	30.68;	Score 390;	DB 19;	Length 416;
Best Local Similarity	97.68;	Pred. No. 2.18e-247;		
Matches . 405;	Conservative	0;	Mismatches 8;	Indels 2;
				Gaps 2

D	b	1	gacccatgctccgaagaccattgtgagtgctttacacacccctgtgaaatcgtccgc	60
Q	y	851	gattcctatctctgccagagaccattgttgatgcttttaccacacctgtgaattgccttc	910
D	b	61	acagttacacaacagctgtgataagtgtgttcctcacacagaaaaactaatgatttcct	120
Q	y	911	acagttacacaacagcgtggaatagtagtggttcacftacacagaaaaactaatgatttcct	970
D	b	121	ggaatcacnctgatatctgtcatatcgtgtgttcattcgtgatatagaagtaaatagt	180
Q	y	971	ggatcacnactatatttgcctaatcgtgttcattcgtgatgtattagagtaaaagtgt	1030
D	b	181	agcctttcaagctttaaatgttgaactatctcaacaaagtaaatctgcgtgact	240
Q	y	1031	agcctttcnaagcttttaaatgtttgaactcatcttaactaaagtaaatctgcgtgact	1090
D	b	241	aatccaatatctccagaaatgtttatcccatctaaagcaattttcatctccaactaagata	300
Q	y	1091	aatccaatatctccagaaatgtttatcccatctaaagcaattttcatctccaactaagata	1150
D	b	301	cttttagcacatgcttaataatacaaaagcgttgtcaatttggaaagtactgtgaaatagat	360
Q	y	1151	ctttttacacatcctttaaataatcaaaaagcagttgtctatttggaaagtaactgtgaaatagat	1210
D	b	361	gttgcaagccgncacactatgntgtatattgttttnccttatgttaggagataaa	415
Q	y	1211	gttgcaagccg-gagacacatatttgatattatattgtt-acctattgattggaataataaa	1263

RESULT 5 standard: cDNA: 91 BP.  
 ID 051746;  
 AC 051746;  
 DT 31-MAY-1994 (first entry)  
 DE Oligonucleotide probe MK14-A  
 KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;  
 ss.  
 OS Synthetic.  
 PN EP-571911-A.  
 PD 01-DEC-1993.  
 PE 24-MAY-1993: 108325.  
 PR 26-MAY-1992: US-889651.  
 PA (BECT ) BECTON DICKINSON CO.  
 PI Shank DD, Spears PA;  
 DR WPI: 93-378844/48.  
 PT New oligo:nucleotide probes specific for Mycobacteria - used for  
 PT detection and amplification of Mycobacteria nucleic acid in  
 PT samples  
 PS Claim 3; Page 14; 23pp: English.  
 CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14  
 CC (051735). It hybridized to all spp. of mycobacteria tested, but  
 CC cross reacted to a few non-mycobacterial spp. The probe may  
 CC be useful as an initial screen for mycobacterial infection.  
 CC See also 051735-45 and 051747-59.  
 SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match	3.58;	Score 44;	DB 9;	Length 91;	..
Best Local Similarity	5.68;	Pred. No. 2.95e-11;			
Matches	3;	Conservative	46;	Mismatches	5; Indels 0; Gaps 0

D<sub>b</sub>      7    ggcgssshsyvvvhvshhsnhvhhvsvvvhhvhhvhhvhyvsy    60  
         ||| :::::::::::::::::::: : : :::::::::::::::  
Q<sub>y</sub>     371   GGTGCATTCCTCAGACTTGGGAAGACCCAGG6CACAAATGATAAACATCACTGGC   424

ID	Standard; DNA; 204 BP.
N81164	

DT 08-NOV-1990 (first entry)  
DE Base substituted E.coli beta-galactosidase alpha-fragment.  
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss  
KW Escherichia coli

FH	Key	Location/Qualifiers
1		

```
      09-08-76    / *tag= a
```

FT	primer bind	/function
FT	187	204

FT	FB-385123-A	/*tag=	b
DN			

PD 05-MAY-1988.

PR 03-APR-1987; US-034819.

PA (SUSO) SUOMEN SOKERI OY.  
Lehtovaara P., Knowles J., Koivuola A., Bamford J., Reinikainen T.

DR WPI; 88-279927/40.

PT by prepn of single stranded template, annealing a primer, elongation,

PS Disclosure; p; English.

Random point mutations were introduced into the alpha fragment of E. coli beta-galactosidase. The wild type sequence was obtained as

single stranded template and an oligonucleotide was hybridised to it to generate a point of RNA molecules which terminated at 31)

CC possible nucleotide positions within a specified region. The

reverse transcriptase. Nucleotides are misincorporated by the

transcriptase and the molecules are completed to forms that can be amplified and then expressed in a suitable post-vector system.

CC The sequence covers all 176 diff't base substitutions, most of which  
CC occurred singularly in any given mutant

CC See also P80575.

Sequence	204 BF	21 A	4 / C	1 / G	11 I	108 UCNEIS
30						

```
Query Match , 3.38; Score 42; DB 1; Length 204;
Best Local Similarity 11.08; Pred NO 4 25e-10;
```

Matches	11;	Conservative	53;	Mismatches	36;	Indels	0;	Gaps	0;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

Db 92 hhyr'mrbnvyr'dynr'sdaa'awycyrrsvkycdcynachddhyvbybbvynvnhnnc 151

282 CTTAAACCCTATTAAACAAGATGTGAAAAAAGGAAAACTTCGCTATGTGCGAATTGT 341

153 nccobnrbvrbhbnbrnswrwhdarddvhcwrhc 101

[illegible]

xy . 342 1CCCC1AAAAGAA1AA1C1CGAAAC1A1GG1GCCCC 301

**Результат**

ID Q51746 standard; cDNA; 91 BP.

DT 31-MAY-1994 (first entry)

DE Oligonucleotide probe MK14-A  
KW Oligonucleotide: DNA probe: mycobacteria: disease diagnosis::

synthetic  
SS,  
KW

PN EP-571911-A.  
PD 01-DEC-1002

PF 24-MAY-1993; 108325.

PA (BECT ) BECTON DICKINSON CO.

F1 Shanks LJ, Spears PA;  
 WPI; 93-37884/48.  
 DR

PT New oligo: nucleotide probes specific for Mycobacterium - used for detection and amplification of Mycobacteria nucleic acid in

[illegible]



```
PS Example 5: Page 35: 71pp: English.
CC A method for treating airway disease in a subject has been produced,
CC which involves the topical administration of an essentially adenosine
CC free antisense oligonucleotide (ON) to the airway epithelium of the
CC subject. The present sequence is an antisense oligonucleotide specific
CC for the human RANP5, targeted at the initiation codon. The method can
CC be used to treat airway diseases such as cystic fibrosis, asthma,
CC chronic obstructive pulmonary disease, bronchitis and other airway
CC diseases characterised by an inflammatory response. By eliminating
CC adenosine from the antisense ON, its liberation upon antisense
CC degradation is prevented, thereby preventing adenosine-induced
CC bronchoconstriction in patients with hyper-reactive airways.
SQ Sequence 162 BP; 0 A; 29 C; 74 G; 22 T;

Query Match 2.8%; Score 36; DB 32; Length 162;
Best Local Similarity 42.9%; Pred. No. 1.04e-06;
Matches 54; Conservative 29; Mismatches 42; Indels 1; Gaps 1;

Db 14 tgggagcgcbtgcgttcgacgcbgagtgatcgcggbtblttggaggcgb 73
| | | | | : | : | : | | | | | : | | | | | : | | | |
Qy 9 TNGGGCTCCTTCCTTGTGACGTGCGCCGCGGCTGCTGCTGTGCAAGCGGC 68
| | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db 74 tgcgagcgcbtgcgcbgagcgctgttcggttgcgcbgagcgccgagbgpact 132
| | | | | : | | | | | : | | | | | : | | | | | : | | | |
Qy 69 GGCGGCAAGACTCGGCACATATGACGGCTTGACGACCAGAGAGCGCCGCCCTTTCT 128
| | | | | : | | | | | : | | | | | : | | | | | : | | | |
Db 133 tcbtg 138
| | | |
Qy 129 CCTGTG 134

RESULT 12
ID V44650 standard: DNA; 91 BP.
AC V44650;
DT 06-OCT-1998 (first entry)
DE Mammalian DNA replication origin consensus sequence, uniorconsensus;
KW DNA replication origin; human; mammal; alphaconsensus; uniorconsensus;
KW anti-gene; DNA replication inhibitor; shuttle vector construct creation;
KM gene therapy; ss.
PN Mammalia.
OS WO9827200-A2.
PD 25-JUN-1998.
PF 12-DEC-1997; CA0972.
PR 21-MAY-1997; US-047322.
PA 16-DEC-1996; US-033374.
PR (UYMC-) UNIV MCSTILL.
PI COSCONS NH, Nielsen TO, Price GB, Zannis-Hadjopoulos M;
PT WPI: 98-562770/31.
PT Human or mammalian origin of replication consensus sequences - for
PT inhibiting DNA replication, for controlling initiation of
PT replication, maintaining circular plasmids and in assembly of human
PT artificial chromosomes
PS Claim 1: Page 42: 54pp: English.
SC This sequence represents a human or mammalian DNA replication origin
CC consensus sequences of the invention, designated uniorconsensus.
CC Administration of the consensus sequence or an anti-gene comprising a
CC double stranded copy of the consensus) is used to inhibit DNA replication
CC in vivo or in vitro. The consensus sequences can also be inserted into an
CC expression vector, used subsequently for in vitro transfection of
CC mammalian cells, to control initiation of DNA replication. They can also
CC be used used to maintain circular plasmids that are capable of.
CC semi-conservative replication in proliferating mammalian cells, or
CC inserted into mammalian or human artificial chromosome vectors for gene
CC therapy. Particularly, they are used to create shuttle vector constructs
CC for defining the essential mammalian elements required for maintenance of
CC chromosomal function. The consensus sequence can be combined with cloned
CC human telomeres and large centromeric blocks for assembly of human
CC artificial chromosomes and maintained as bacterial plasmids, circular or
CC linear, large or small yeast artificial chromosomes (YACS) or as episomal
CC elements.
CC Sequence 91 BP; 15 A; 1 C; 4 G; 7 T;

Query Match 2.7%; Score 34; DB 46; Length 91;
```

[illegible][illegible]

US-09-415-540-2.rng

[illegible]

Search completed: Mon Dec 27 13:07:45 1999  
Job time : 830 secs.



\*\*\*\*\*  
Molecular  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

\*\*\*\*\*  
MSPRch\_n n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Mon Dec 27 13:08:02 1999; Maspar time 92.62 Seconds  
1190.364 Million cell updates/sec  
Tabular output not generated.

Title: >US-09-415-540-2  
Description: (1-1275) from US09415540.seq  
Perfect Score: 1274  
N.A. Sequence: 1. CAAGAGGTGCGGCTCTCT.....GAATTAATTAATTTGCTG 1275  
Comp: GTTCCACANCCCGAGAGA.....CTTATTTAATTAACGAC

Scoring table: TABLE default  
Gap 6

Match STD : Dbase 0; Query 0

Searched: 165362 seqs, 43234748 bases x 2

Post-processing: Minimum Match 08  
Listing first 45 summaries

Database: n-issued  
1:5A\_COMB 2:5B\_COMB 3:5C\_COMB 4:PCT9\_COMB 5:backfiles1

Statistics: Mean 8.783; Variance 4.614; scale 1.904

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	1274	100.0	1275	3	US-08-741-	Sequence 2, Applicatio	0.00e+00
2	80	6.3	7218	2	US-08-232-	Sequence 14, Applicati	5.75e+40
3	34	2.7	965	3	US-08-388-	Sequence 22, Applicati	1.04e-07
4	28	2.2	215	1	US-08-238-	Sequence 5, Applicatio	4.80e-04
5	28	2.2	965	3	US-08-388-	Sequence 22, Applicati	4.80e-04
6	27	2.1	215	1	US-08-238-	Sequence 5, Applicatio	1.85e-03
7	24	1.9	65	1	US-08-471-	Sequence 145, Applicat	9.22e-02
8	24	1.9	68	1	US-08-471-	Sequence 143, Applicat	9.22e-02
9	24	1.9	74	4	PCT-US95-1	Sequence 100, Applicat	9.22e-02
10	24	1.9	75	4	PCT-US95-1	Sequence 99, Applicati	9.22e-02
11	24	1.9	75	4	PCT-US95-1	Sequence 99, Applicati	9.22e-02
12	24	1.9	81	4	PCT-US95-1	Sequence 97, Applicati	9.22e-02
13	24	1.9	82	4	PCT-US95-1	Sequence 97, Applicati	9.22e-02
14	24	1.9	82	4	PCT-US95-1	Sequence 97, Applicati	9.22e-02
15	23	1.8	74	4	PCT-US95-1	Sequence 94, Applicati	3.24e-01
16	23	1.8	74	4	PCT-US95-1	Sequence 94, Applicati	3.24e-01
17	23	1.8	81	4	PCT-US95-1	Sequence 92, Applicati	3.24e-01
18	23	1.8	81	4	PCT-US95-1	Sequence 92, Applicati	3.24e-01
19	23	1.8	92	3	US-08-353-	Sequence 16, Applicati	3.24e-01
20	23	1.8	396	4	PCT-US95-1	Sequence 27, Applicati	3.24e-01

RESULT	ID	US-08-741-437-2 STANDARD, DNA; UNC; 1275 BP.	ALIGNMENTS
AC	xxxxxx		
DE	Sequence 2, Application US/08741437		
CC	Sequence 2, Application US/08741437		
CC	Patent No. 5843665		
CC	GENERAL INFORMATION:		
CC	APPLICANT: Hawkins, Phillip R.		
CC	APPLICANT: Hillman, Jennifer L.		
CC	TITLE OF INVENTION: A NOVEL HUMAN PYROPHOSPHATASE		
CC	NUMBER OF SEQUENCES: 15		
CC	CORRESPONDENCE ADDRESS:		
CC	ADDRESSEE: Incyte Pharmaceuticals, Inc.		
CC	STREET: 3174 Porter Drive		
CC	CITY: Palo Alto		
CC	STATE: California		
CC	COUNTRY: USA		
CC	ZIP: 94304		
CC	COMPUTER READABLE FORM:		
CC	MEDIUM TYPE: Diskette		
CC	COMPUTER: IBM Compatible		
CC	OPERATING SYSTEM: DOS		
CC	SOFTWARE: FastSeq Version 1.5		
CC	CURRENT APPLICATION DATA:		
CC	APPLICATION NUMBER: US/08/741,437		
CC	FILING DATE: Filed Herewith		
CC	Prior Application Data:		
CC	APPLICATION NUMBER:		
CC	FILING DATE:		
CC	ATTORNEY/AGENT INFORMATION:		
CC	NAME: Billings, Lucy J.		
CC	REGISTRATION NUMBER: 36,749		
CC	REFERENCE/DOCKET NUMBER: PF-0148 US		
CC	TELECOMMUNICATION INFORMATION:		
CC	TELEPHONE: (415) 855-0555		
CC	TELEFAX: (415) 845-4166		
CC	INFORMATION FOR SEQ ID NO: 2:		
CC	SEQUENCE CHARACTERISTICS:		
CC	LENGTH: 1275 base pairs		
CC	TYPE: nucleic acid		
CC	STRANDEDNESS: single		
CC	TOPOLOGY: linear		
CC	MOLECULE TYPE: cDNA		

IMMEDIATE SOURCE:  
LIBRARY:  
CLONE: Consensus  
SEQUENCE 1275 BP; 394 A; 235 C; 294 G; 351 T; 1 OTHER.

Query Match 100.0%; Score 1274; DB 3; Length 1275;  
Best Local Similarity 100.0%; Pred. No. 0.00e+00;  
Matches 1275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 CAAGAGGTTNGGGCTCTCTCTTGTGCACTCGGCGCGCGCTGCGGGCTGTGCTGTG 60  
QY 1 CAAGAGGTTNGGGCTCTCTCTTGTGCACTCGGCGCGCGCTGCGGGCTGTGCTGTG 60  
DB 61 GAGAGGGGGGGGGGAGGAGTCCGGCACTATGAGCGGCTTACAGCCAGAGAGCGCGCC 120  
QY 61 GAGAGGGGGGGGGGAGGAGTCCGGCACTATGAGCGGCTTACAGCCAGAGAGCGCGCC 120  
DB 121 GCCCTTCTCCCTGGAGTCCGAGTCTTCTCTCAAAAATGAGAAGGACATTAATCTCC 180  
QY 121 GCCCTTCTCCCTGGAGTCCGAGTCTTCTCTCAAAAATGAGAAGGACATTAATCTCC 180  
DB 181 ATTTCATGATATCCAAATTATGACAGATAGAGATGTGTTACATGATGTTGAAGTACC 240  
QY 181 ATTTCATGATATCCAAATTATGACAGATAGAGATGTGTTACATGATGTTGAAGTACC 240  
DB 241 ACGTGCTATATGCAAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACA 300  
QY 241 ACGTGCTATATGCAAAAATGAGATGCTACAAAGACCCCTTTAAACCTATTAAACA 300  
DB 301 AGATGTGAAAAAGGAAAACTTCGTATGTTGCGAATTTGTTCCGTAATAAGATATAT 360  
QY 301 AGATGTGAAAAAGGAAAACTTCGTATGTTGCGAATTTGTTCCGTAATAAGATATAT 360  
DB 361 CTGAGACTATGTGCTCCTCCTCAACACTTGGGAAGACCCAGGGCCACATGAATAATAC 420  
QY 361 CTGAGACTATGTGCTCCTCCTCAACACTTGGGAAGACCCAGGGCCACATGAATAATAC 420  
DB 421 TGGCTGTGTGTGTCAGTATGACCAATGATGTTGTAATGGAAGCAAGTATGTC 480  
QY 421 TGGCTGTGTGTGTCAGTATGACCAATGATGTTGTAATGGAAGCAAGTATGTC 480  
DB 481 AAGAGGTAAATTAATGCGTGTAAGTTCTAGCATATGCGTATGATGAGAGGGA 540  
QY 481 AAGAGGTAAATTAATGCGTGTAAGTTCTAGCATATGCGTATGATGAGAGGGA 540  
DB 541 AACCAGCTGGAAGTCAATGCCATTAAATGATGATGATGATGATGATGATGATGATG 600  
QY 541 AACCAGCTGGAAGTCAATGCCATTAAATGATGATGATGATGATGATGATGATGATG 600  
DB 601 TATCAATGATGCAAAAGCGTGAACCTGCGTACTTAAAGCTACTGTGAGTGTG 660  
QY 601 TATCAATGATGCAAAAGCGTGAACCTGCGTACTTAAAGCTACTGTGAGTGTG 660  
DB 661 AAGGATTAAGGTTCTGTATGTAAGAAACAGAAATAGATTGCGTTAATGAGATTAA 720  
QY 661 AAGGATTAAGGTTCTGTATGTAAGAAACAGAAATAGATTGCGTTAATGAGATTAA 720  
DB 721 AGATTAAGGACTTGGCAATGATTAATTAAGACACTATGACCATGGAAGCAATTA 780  
QY 721 AGATTAAGGACTTGGCAATGATTAATTAAGACACTATGACCATGGAAGCAATTA 780  
DB 781 GACTTAAGAAAGATGGAAGAAAGCAATGATGATGATGATGATGATGATGATGATG 840  
QY 781 GACTTAAGAAAGATGGAAGAAAGCAATGATGATGATGATGATGATGATGATGATG 840  
DB 841 CTGAAGTGTATCTGTATGCTGACAGAGCAATGATGATGATGATGATGATGATGATG 900  
QY 841 CTGAAGTGTATCTGTATGCTGACAGAGCAATGATGATGATGATGATGATGATGATG 900  
DB 901 ATCTGCTGCAACAGTACCAAGAGAGTGAATAGTGTTCATCCAGAGAAAAAATTA 960  
QY 901 ATCTGCTGCAACAGTACCAAGAGAGTGAATAGTGTTCATCCAGAGAAAAAATTA 960

DB 961 AGATTCTCTGGAATACAGCGTATGTCATACATGCTGTCATCTGATGATAGAG 1020  
QY 961 AGATTCTCTGGAATACAGCGTATGTCATACATGCTGTCATCTGATGATAGAG 1020  
DB 1021 TAAAGTGTAGCTTTTCAAGCTTTAAATTTGTGAACATCATCACTAAAGTAAATTC 1080  
QY 1021 TAAAGTGTAGCTTTTCAAGCTTTAAATTTGTGAACATCATCACTAAAGTAAATTC 1080  
DB 1081 TGTCTGACTATCCAAATACCTCAGAAATGTTATCATCTAAAGCATTTTTCATATCA 1140  
QY 1081 TGTCTGACTATCCAAATACCTCAGAAATGTTATCATCTAAAGCATTTTTCATATCA 1140  
DB 1141 ACTAGATTAATCTTTAGACATGCTTAATATCAAGAGAGTGTATTTGGAAGTCACTT 1200  
QY 1141 ACTAGATTAATCTTTAGACATGCTTAATATCAAGAGAGTGTATTTGGAAGTCACTT 1200  
DB 1201 GTGATTAATGTGCAAGGGGAGACATATGATGATGATGATGATGATGATGATGATG 1260  
QY 1201 GTGATTAATGTGCAAGGGGAGACATATGATGATGATGATGATGATGATGATGATG 1260  
DB 1261 AAAATTATTTGCTG 1275  
QY 1261 AAAATTATTTGCTG 1275

RESULT 2  
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.  
AC xxxxxx

DE Sequence 14, Application US/08232463  
CC Sequence 14, Application US/08232463  
CC Patent No. 5670367

CC GENERAL INFORMATION:  
CC APPLICANT: DORNER, F.  
CC APPLICANT: SCHEIFLINGER, F.  
CC APPLICANT: FALKNER, F. G.  
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
CC NUMBER OF SEQUENCES: 52

CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Foley & Lardner  
CC STREET: 1800 Diagonal Road, Suite 500  
CC CITY: Alexandria  
CC STATE: VA  
CC COUNTRY: USA

CC ZIP: 22313-0299

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patentin Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/232,463

CC FILING DATE:

CC CLASSIFICATION: 435

CC PRIORITY APPLICATION DATA:  
CC APPLICATION NUMBER: US/07/935,313

CC FILING DATE:

CC APPLICATION NUMBER: EP 91 114 300.6

CC FILING DATE: 26-AUG-1991

CC ATTORNEY/AGENT INFORMATION:  
CC NAME: BENT, Stephen A.

CC REGISTRATION NUMBER: 29,768

CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU

CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (703)836-9300

CC TELEFAX: (703)683-4109

CC TELEX: 899149

CC INFORMATION FOR SEQ ID NO: 14:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 7218 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC	IMMEDIATE SOURCE:
CC	CLONE: PTZgpt-F1S
CC	SEQUENCE 7218 BP; 1944 A; 1491 C; 1486 G; 1929 T; 368 OTHER
SQ	

Query Match	6.3%	Score 80;	DB 2;	Length 7218;
Best Local Similarity	0.8%;	Pred. No. 5.75e-40;		
Matches	3;	Conservative 217;	Mismatches 140;	Indels 0;
			Gaps 0;	

[illegible]

RESULT	3
ID	US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP

DE	Sequence 22, Application	US/08388672A
CC	Sequence 22, Application	US/08388672A
CC	Patent No. 5795961	

CC GENERAL INFORMATION:  
CC APPLICANT: Wallace, T. Paul  
CC APPLICANT: Harris, William J.  
CC APPLICANT: Carr, Frank J.  
CC APPLICANT: Old, Lloyd J.  
CC APPLICANT: Welt, Sydney  
CC APPLICANT: Kilnauir, Kunio  
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
CC TITLE OF INVENTION: Antibodies  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:

CC STATE: New York  
CC COUNTRY: U.S.A.  
CC ZIP: 10022  
CC  
CC COMPUTER READABLE FORM:  
CC  
CC MEDIUM TYPE: Floppy disk  
CC  
CC COMPUTER: IBM PC compatible  
CC  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC  
CC SOFTWARE: Patentin Release #1.0, Version #1.30  
CC  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/388,672A  
CC  
CC FILING DATE: 14-FEB-1995

```

CC INFORMATION FOR SEQ ID NO: 22:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 965 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: DNA (genomic)
CC SEQUENCE 965 BP: 192 A; 170 C; 226 G; 200 T; 177 OTHER
SD

```

Query Match	2.7%;	Score 34;	DB 3;	Length 965;
Best Local Similarity	12.9%;	Pred. No. 1.04e-07;		
Matches	15;	Conservative	62;	Mismatches 37; Indels 2; Gaps 2.

Db 826 YNKGKRYTADTADTSNNSSTVADTADTVYVCRRSRSDSDGDDWGCT -VTYSHYDK 884  
 Qy 8 TTNGGGCTCTCTCTCTTGTACGTGCGGGCGCGCGGGCGGCGGCTCTGTGTGACACGG 67  
 Db 885 MTSSSSASVGRVYTCRSTTHGNGNTIYWKAKXRVSNRSGVSSGSGCDY 940  
 Qy 68 CGCGGG -CAGACATCTCGGCACTATGACGCGCTTGACGACGAGGAGCGCGCGCGC 122

RESULT	4
ID	US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.

DE	Sequence 5, Application US/08238163
CC	Sequence 5, Application US/08238163
CC	Patent No. 5569830
CC	GENERAL INFORMATION.

CC GENERAL INFORMATION:  
CC APPLICANT: BENNETT, Alan  
CC APPLICANT: LABAYTCH, John M.  
CC APPLICANT: POWELL, Ann  
CC APPLICANT: STOTZ, Henrik  
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL  
CC TITLE OF INVENTION: POLYGLACTURONASES AND THEIR USE TO CONTROL FUNGAL DISEASE  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESSES:

CC ADDRESS: Townsend and Townsend Knourie and Crew  
CC STREET: Stewart Street Tower, One Market Plaza  
CC City: San Francisco  
CC STATE: California  
CC COUNTRY: US

CC ZIP: 94105-1493  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC COMPUTER: IBM PC compatible  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent Release #1.0.  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/238,163  
CC FILING DATE: 03-MAY-1994  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:

```

CC NAME:., Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/POCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc_feature
CC LOCATION: 1..215
CC OTHER INFORMATION: /standard_name="Deduced amino acid"
CC OTHER INFORMATION: sequence of PCpI from bean."
SO SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

```

SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.



Query Match 2.2% Score 28; DB 1; Length 215;  
Best Local Similarity 13.3%; Pred. No. 4,80e-04;  
Matches 20; Conservative 62; Mismatches 67; Indels 1; Gaps 1;

Db 1 MTNMTSSSVSTRASNDKAKKDGNT-TSSMTTDCNRWGCDDDTTRVNNDSHN 59  
22 CTTGTCACTGCGCGCGCGCTGCTGTGTGCGAGCGCGCGCGAGACTC 81  
OY 60 KYSSANYGNNVCAKTHYTHNVSADSKVTDSYNASGTSSTSSNGTDCNRSGADS 119  
OY 82 CGGACATATGAGCGGCTTACAGACGAGAGAGCGCGCGCTTCTCCGAGATACG 141  
Db 120 YGSKRTAMTSNRRTGKTANNAVDSRNGDA 149  
OY 142 AGCTCTCTCAAAATGAGAAAGCACAATA 171

RESULT 5  
ID US-08-388-672A-22 STANDARD; DNA; UNC; 965 BP.  
AC xxxxxx

DE Sequence 22, Application US/08388672A  
CC Sequence 22, Application US/08388672A  
CC Patent No. 5795961  
CC GENERAL INFORMATION:  
CC APPLICANT: Wallace, T. Paul  
CC APPLICANT: Harris, William J.  
CC APPLICANT: Carr, Frank J.  
CC APPLICANT: Old, Lloyd J.  
CC APPLICANT: Well, Sydney  
CC APPLICANT: Kitamura, Kunio  
CC TITLE OF INVENTION: Recombinant Human Anti-Lewis B  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESS: Felfe and Lynch  
CC STREET: 805 Third Avenue  
CC CITY: New York  
CC STATE: New York  
CC COUNTRY: U.S.A.  
CC ZIP: 10022  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.30  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/388,672A  
CC FILING DATE: 14-FEB-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Hanson, No. 5795961man D.  
CC REGISTRATION NUMBER: 30,946  
CC REFERENCE/DOCKET NUMBER: LND 5409  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 212-688-9200  
CC TELEFAX: 212-638-3884  
CC INFORMATION FOR SEQ. ID NO: 22:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 965 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: unknown  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: DNA (genomic)  
SQ SEQUENCE 965 BP; 192 A; 170 C; 226 G; 200 T; 177 OTHER.

Query Match 2.2% Score 28; DB 3; Length 965;  
Best Local Similarity 13.6%; Pred. No. 4,80e-04;  
Matches 25; Conservative 83; Mismatches 74; Indels 2; Gaps 2;

Db 780 KGRHVHVGVRSTASDYTTSTYGVWVGRGWGYYGGYTYNCKRGRTVMDATS 839

Cp 1003 TGAACAGATGTAGCAATATACCTTGATTC-ACAGAAATCTCATAGTTTCTGCT 945  
Db 840 SNNRSYTAADTAVYCVGRSVDSDGDTTAVTSSHVYKDMTSSSASVGRVTT 899  
Cp 944 GATGAGACACTATATCACGCTGTGTGCTGACGAGCAGATTCACAGGGTGTGCTA 885  
Db 900 CRSTTGNNTTYWYKRAKRYVSNSSVS-R-SGSSGTDYTTSSDATTYCGTHARTG 958  
Cp 884 AAGCATCACAAATGCTCTGCGACATCAGATCATTGAAGGGCTCTCAGACAAG 825  
Db 959 TKVX 962  
Cp 824 TTGT 821

RESULT 6  
ID US-08-238-163-5 STANDARD; DNA; UNC; 215 BP.  
AC xxxxxx

DE Sequence 5, Application US/08238163  
CC Sequence 5, Application US/08238163  
CC Patent No. 5569830  
CC GENERAL INFORMATION:  
CC APPLICANT: BENNETT, Alan  
CC APPLICANT: LABAVITICH, John M.  
CC APPLICANT: POWELL, Ann  
CC APPLICANT: STORZ, Henrik  
CC TITLE OF INVENTION: PLANT INHIBITORS OF FUNGAL  
CC NUMBER OF SEQUENCES: 24  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESS: Townsend and Townsend Kourie and Crew  
CC STREET: Stuart Street Tower, One Market Plaza  
CC CITY: San Francisco  
CC STATE: California  
CC COUNTRY: US  
CC ZIP: 94105-1493  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Floppy disk  
CC OPERATING SYSTEM: PC-DOS/MS-DOS  
CC SOFTWARE: Patent In Release #1.0, Version #1.25  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/238,163  
CC FILING DATE: 03-MAY-1994  
CC CLASSIFICATION: 800  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Bastian, Kevin L.  
CC REGISTRATION NUMBER: 34,774  
CC REFERENCE/DOCKET NUMBER: 2307E-540  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (415) 543-9600  
CC TELEFAX: (415) 543-5043  
CC INFORMATION FOR SEQ. ID NO: 5:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 215 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: unknown  
CC MOLECULE TYPE: protein  
CC FEATURE:  
CC NAME/KEY: misc feature  
CC LOCATION: 1..215  
CC OTHER INFORMATION: /standard\_name="Deduced amino acid  
CC OTHER INFORMATION: sequence of p51p from bean."  
SQ SEQUENCE 215 BP; 15 A; 8 C; 25 G; 26 T; 141 OTHER.

Query Match 2.1% Score 27; DB 1; Length 215;  
Best Local Similarity 15.0%; Pred. No. 1,85e-03;  
Matches 25; Conservative 65; Mismatches 75; Indels 2; Gaps 2;

Db 39 RTWGVCDTDT-YYVNNDSGNKYSANYNNGANNVGAATHTYTHNVSADSKTYTDS 97



CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mistrock, S. Leslie  
CC REGISTRATION NUMBER: 18,872  
CC REFERENCE/DOCKET NUMBER: 1101-196-228  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 790-9090  
CC TELEFAX: (212) 869-9741/8864  
CC TELETYPE: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 100:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 74 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 74 BP: 6 A; 6 C; 1 G; 1 T; 60 OTHER.  
SO

Query Match 1.9%; Score 24; DB 4; Length 74;  
Best Local Similarity 8.7%; Pred. No. 9.22e-02;  
Matches 6; Conservative 20; Mismatches 43; Indels 0; Gaps 0;

Db 6 VNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN 65  
Cc 74 GCCCGCCGCCCTGCCACAGACGACCACCCCGCGCGCGACTGACACAGAGAGAG 15  
Db 66 ACCACACACC 74  
Cc 14 CCCCNAACC 6  
Cp

RESULT 10  
ID PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.  
AC xxxxxx  
DT Sequence 99, Application PC/TUS9511934  
DE Sequence 99, Application PC/TUS9511934  
CC GENERAL INFORMATION:  
CC APPLICANT: Cytogen Corporation  
CC TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From  
CC NUMBER OF SEQUENCES: 103  
CC CURRENT APPLICATION DATA:  
CC FILING DATE: 20-SEP-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mistrock, S. Leslie  
CC REGISTRATION NUMBER: 18,872  
CC REFERENCE/DOCKET NUMBER: 1101-196-228  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 790-9090  
CC TELEFAX: (212) 869-9741/8864  
CC TELETYPE: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 99:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 75 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 75 BP: 1 A; 1 C; 7 G; 5 T; 61 OTHER.  
SO

Query Match 1.9%; Score 24; DB 4; Length 75;  
Best Local Similarity 7.7%; Pred. No. 9.22e-02;  
Matches 5; Conservative 20; Mismatches 40; Indels 0; Gaps 0;

Db 4 AGNNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBN 63  
Cc 3 AGAGCTTNGGGGCTCTCTGTCATGCGCGCGCGCTGGGCTGTGCTGTGCGC 62  
Db 64 NNBGG 68  
Cc 63 AGCGG 67  
Cp

RESULT 11  
ID PCT-US95-11934-99 STANDARD; DNA; UNC; 75 BP.  
AC xxxxxx  
DT Sequence 99, Application PC/TUS9511934  
DE Sequence 99, Application PC/TUS9511934  
CC GENERAL INFORMATION:  
CC APPLICANT: Cytogen Corporation  
CC TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From  
CC NUMBER OF SEQUENCES: 103  
CC CURRENT APPLICATION DATA:  
CC FILING DATE: 20-SEP-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mistrock, S. Leslie  
CC REGISTRATION NUMBER: 18,872  
CC REFERENCE/DOCKET NUMBER: 1101-196-228  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 790-9090  
CC TELEFAX: (212) 869-9741/8864  
CC TELETYPE: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 99:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 75 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 75 BP: 1 A; 1 C; 7 G; 5 T; 61 OTHER.  
SO

Query Match 1.9%; Score 24; DB 4; Length 75;  
Best Local Similarity 7.5%; Pred. No. 9.22e-02;  
Matches 5; Conservative 20; Mismatches 42; Indels 0; Gaps 0;

Db 5 GNNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBNBN 64  
Cc 442 GTCATTGTCACCAACAGCAGTATGTTATCATGTCGCCCTGGCTTCCCAAGTCTG 383  
Db 65 NNBGTTG 71  
Cc 382 AGGATG 376  
Cp

RESULT 12  
ID PCT-US95-11934-98 STANDARD; DNA; UNC; 81 BP.  
AC xxxxxx  
DT Sequence 99, Application PC/TUS9511934  
DE Sequence 99, Application PC/TUS9511934  
CC GENERAL INFORMATION:  
CC APPLICANT: Cytogen Corporation  
CC TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From  
CC NUMBER OF SEQUENCES: 103  
CC CURRENT APPLICATION DATA:  
CC FILING DATE: 20-SEP-1995  
CC CLASSIFICATION:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Mistrock, S. Leslie  
CC REGISTRATION NUMBER: 18,872  
CC REFERENCE/DOCKET NUMBER: 1101-196-228  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: (212) 790-9090  
CC TELEFAX: (212) 869-9741/8864  
CC TELETYPE: 66141 PENNIE  
CC INFORMATION FOR SEQ ID NO: 99:  
CC SEQUENCE CHARACTERISTICS:  
CC LENGTH: 75 base pairs  
CC TYPE: nucleic acid  
CC STRANDEDNESS: single  
CC TOPOLOGY: linear  
CC MOLECULE TYPE: DNA (genomic)  
CC SEQUENCE 81 BP: 1 A; 1 C; 7 G; 5 T; 61 OTHER.  
SO

```

AC      xxxxxx
DT      Sequence 98, Application PC/TUS9511934
DE      Sequence 98, Application PC/TUS9511934
CC      GENERAL INFORMATION:
CC      APPLICANT: Cytogen Corporation
CC      TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC      TITLE OF INVENTION: Peptide Libraries
CC      NUMBER OF SEQUENCES: 103
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Pennie & Edmonds
CC      STREET: 1155 Avenue of the Americas
CC      CITY: New York
CC      STATE: New York
CC      COUNTRY: USA
CC      ZIP: 10036
CC      COMPUTER READABLE FORM:
CC      MEDIUM TYPE: Floppy disk
CC      COMPUTER: IBM PC compatible
CC      OPERATING SYSTEM: PC-DOS/MS-DOS
CC      SOFTWARE: Patentln Release #1.0, Version #1.30
CC      CURRENT APPLICATION DATA:
CC      APPLICATION NUMBER: PCT/US95/11934
CC      FILING DATE: 20-SEP-1995
CC      CLASSIFICATION:
CC      ATTORNEY/AGENT INFORMATION:
CC      NAME: Mistrock, S. Leslie
CC      REGISTRATION NUMBER: 18,872
CC      REFERENCE/DOCKET NUMBER: 1101-196-228
CC      TELECOMMUNICATION INFORMATION:
CC      TELEPHONE: (212) 790-9090
CC      TELEFAX: (212) 869-9741/8864
CC      TELE:
CC      INFORMATION FOR SEQ ID NO: 98:
CC      SEQUENCE CHARACTERISTICS:
CC      LENGTH: 81 base pairs
CC      TYPE: nucleic acid
CC      STRANDEDNESS: single
CC      TOPOLOGY: linear
CC      MOLECULE TYPE: DNA (genomic)
CC      SEQUENCE 81 BP; 6 A; 6 C; 4 G; 5 T; 60 OTHER.
SQ
Query Match      1.9%; Score 24; DB 4; Length 81;
Best Local Similarity 8.7%; Pred. No. 9.22e-02;
Matches      6; Conservative      20; Mismatches 43; Indels 0; Gaps 0;

Db      13 VNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNNVNN
Cp      74 GCGCGCGCGCGTGCACAGACGACACGACCGCGCACGCGCGCGACAGCAGAGAGAG
Db      73 ACCACACACC 81
      11 1111
Cp      14 CCCCNACC 6

RESULT 13
ID      PCT-US95-11934-97 STANDARD; DNA; UNC; 82 BP.
AC      xxxxxx
DE      Sequence 97, Application PC/TUS9511934
CC      Sequence 97, Application PC/TUS9511934
CC      GENERAL INFORMATION:
CC      APPLICANT: Cytogen Corporation
CC      TITLE OF INVENTION: Antigen Binding Peptides (Ablides) From
CC      TITLE OF INVENTION: Peptide Libraries
CC      NUMBER OF SEQUENCES: 103
CC      CORRESPONDENCE ADDRESS:
CC      ADDRESSEE: Pennie & Edmonds
CC      STREET: 1155 Avenue of the Americas
CC      CITY: New York
CC      STATE: New York
CC      COUNTRY: USA
CC      ZIP: 10036

```

[illegible]

